Final Report
2004 Assessment Grant

Dean De Cock

TABLE OF CONTENTS

LIST OF TABLES		iii
LIST OF FIGURES		iv
CHAPTER 1	INTRODUCTION	1
1.1	Placement Scores	1
1.2	ACT Score	1
1.3	High School GPA	2
CHAPTER 2	INTRODUCTORY COURSES	3
2.1	Math 156 (College Algebra)	3
2.2	Math 157 (Trigonometry)	9
2.3	Math 186 (Elementary Functions)	12
2.4	Math 194 (LAS Calculus)	16
2.5	Math 198 (Calculus I)	19
2.6	Introductory Class Summary	22
CHAPTER 3	UPPER LEVEL COURSES	25
3.1	Math 264 (Calculus II)	25
3.2	Advanced Placement (AP)	25
3.3	Data Collection	26
3.4	Policy Change	28
CHAPTER 4	SUMMARY	29
APPENDICES		30

LIST OF TABLES

Table 1	Correlation matrix of grade and explanatory variables for Math 156	3
Table 2	Predicted versus actual grade in Math 156 utilizing the regression model	4
Table 3	Predicted grade distribution in M156 using ordinal logistic regression	7
Table 4	Correlation matrix of grade and explanatory variables for Math 157	9
Table 5	Predicted versus actual grade in Math 157 utilizing the regression model	10
Table 6	Predicted grade distribution in M157 using ordinal logistic regression	11
Table 7	Correlation matrix of grade and explanatory variables for Math 186	12
Table 8	Predicted versus actual grade in Math 186 utilizing the regression model	13
Table 9	Predicted grade distribution in M186 using ordinal logistic regression	15
Table 10	Correlation matrix of grade and explanatory variables for Math 194	16
Table 11	Predicted versus actual grade in Math 194 utilizing the regression model	17
Table 12	Predicted grade distribution in M194 using ordinal logistic regression	18
Table 13	Correlation matrix of grade and explanatory variables for Math 198	19
Table 14	Predicted versus actual grade in Math 198 utilizing the regression model	20
Table 15	Predicted grade distribution in M198 using ordinal logistic regression	21
Table 16	Number of students receiving various grades for Calc II based on AP score	26
Table 17	Percentage of students receiving various grades in Calc II by AP exam score	27

LIST OF FIGURES

Figure 1	Predicted grades for students in Math 156 classified by grade received	5
Figure 2	Predicted grades for students in Math 157 classified by grade received	10
Figure 3	Predicted grades for students in Math 186 classified by grade received	14
Figure 4	Predicted grades for students in Math 194 classified by grade received	17
Figure 5	Predicted grades for students in Math 198 classified by grade received	20
Figure 6	Predicted grades in the introductory courses for a high ability student	22
Figure 7	Predicted grades in the introductory courses for a low ability student	23
Figure 8	Percentage of students receiving various grades in Calc II by AP exam score	27

CHAPTER 1 INTRODUCTION

The purpose of this project was to analyze and improve the system used for placing students into their first mathematics class at Truman. The first part of this report will discuss placement in most common introductory classes of MATH156, MATH186, and MATH198, while the second part will focus on the special issues of MATH263.

The original goal of this study was to create a regression equation using ACT score, math placement test scores, and high school math GPA that would predict a student's success for various math courses. This equation could then be used by the placement officer to assist in placing students into the correct math class at Truman. As data collection for this project began, it was evident various sources of information had different levels of availability.

1.1 Placement Scores

The mathematics placement exam is the primary method used by the department in placing students into their first math class. The test consists of two separate exams with one focusing on basic algebra and the other on trigonometry. The self-administered exam is mailed to all incoming freshman and is taken during the spring semester of their senior year in high school. As this is the primary method of placement, this data is available on almost all students who have attended Truman.

1.2 ACT Score

The ACT scores were easily obtained on almost all individuals that have attended Truman over the last ten years as these scores are currently used for math placement and are part of a student's university record. Rather than using the general ACT composite score, this analysis will utilize the math composite score (ACTM) which is composed of three sub-scores covering Algebra, Geometry and Trigonometry. Although Truman does not currently use the specific math sub-scores for placement, the data was readily

available on most students so these variables were added to the pool of potential explanatory variables.

1.3 High School GPA

Obtaining the GPA for high school mathematics classes was much more problematic than the other variables. The current placement system does incorporate the classes taken and grades received in high school, but only through a review of the transcript at placement time. No records of the student's math GPA or coursework is added to the university record. This information could be obtained via manual reviews of all student application folders but this method would be time prohibitive. Fortuitously, a second method of evaluating high school performance was found. On the ACT application, students are asked to report their high school math GPA. Although this self reported data is not as reliable as transcripts, its use would reduce the time of data collection dramatically.

Initial inquiries with IT indicated that the high school GPA field did exist on our ACT tapes but the information was currently being ignored during downloads. Though many historical tapes had been eliminated, IT was able to obtain information on three years of incoming freshman (2000-2002). This resulted in approximately 2100 students having a full record of explanatory variables (see below) being available for this analysis.

ACTM – Composite math score based on three sub-scores

- Sub 1 – Pre-Algebra/Elementary Algebra
- Sub 2 – Algebra/Coordinate Geometry
- Sub 3 – Plane Geometry/Trigonometry

PT1 – Truman administered exam evaluating algebra skills

PT2 – Truman administered exam evaluating trigonometric skills

SMG – Self reported high school math GPA

NGRADE – Grade received in the first math class taken at Truman

CHAPTER 2 INTRODUCTORY COURSES

2.1 Math 156 (College Algebra)

The investigation began by conducting a simple correlation analysis to see if any of the possible explanatory variables are linearly related to the grade received in the class (or to each other). The strongest linear association with the M156 grade is the self reported high school math GPA. This is not unusual, as it would be expected that students who excelled in high school would also excel in college. It is also evident from Table 1 that PT1 and ACTM have some association with grade, while PT2 shows a minimal relationship to M156 performance.

	Grade	ACTM	Sub 1	Sub 2	Sub 3	PT1	PT2
ACTM	0.261						
Sub 1	0.207	0.708					
Sub 2	0.220	0.687	0.525				
Sub 3	0.208	0.721	0.499	0.501			
PT1	0.321	0.372	0.307	0.390	0.294		
PT2	0.150	0.140	0.059	0.123	0.148	0.411	
SRG	0.334	0.195	0.154	0.123	0.157	0.209	0.103

Table 1 – Correlation matrix of grade and explanatory variables for Math 156

In comparing the correlation of the explanatory variables to themselves we see a strong relationship between ACTM and PT1, which would be expected as they measure similar characteristics via different tests. The score on the two placement exams are also correlated.

As many of the explanatory variables measure similar qualities and are at least partially correlated, a stepwise regression was conducted to determine which were the strongest predictors of performance in M156. The variables chosen, in order of selection, were SMG, PT1, ACTM and PT2. As the p-value for PT2 was substantially greater than

the other three $(p = 0.13)$ it was eliminated from the model. A regression on the remaining variables yields the following resulting equation.

$$\hat{y} = -2.22 + 0.0534ACTM + 0.0702PT1 + 0.613SMG$$

To illustrate the use of the equation two students from the data set were selected (one of high and one of low mathematical ability) and their predicted and actual grades are listed in the table below. The predicted grade for the high student (3.40) can be thought of as the average grade for a large number of students at that ability level. The particular student from the data set achieved an A in M156. The predicted grade for the low student (1.21) indicates that the average student with these abilities will tend to receive a D, while the actual student dropped the class.

Ability	ACTM	PT1	SMG	Predicted	Actual
High	29	23	4.0	3.40	A
Low	17	7	3.3	1.21	W

Table 2 – Predicted versus actual grade in Math 156 utilizing the regression model

Although all of the coefficients in the model are highly significant, the model in general explains only 20% of the variation in the grade received by the student. Attempts to improve the fit of the model via transformations and higher order terms did not result in improvements in fit.

A plot of the grade received in the class versus the predicted performance allows for better evaluation of predictive power of the model. There is a definite trend present in the data with the distribution of predicted grade shifting lower as the actual grade decreases from A to D. What is also apparent is that the predicted grades for students who withdrew or received F's appear more spread out than the other grades. The smaller number of W/F's could explain the lack of mounding for these distributions, but there

still appears to be excessive variation in predicted grades for these groups. Intuitively this might be expected for the students who choose to withdraw from M156, as there will be students who withdraw because they are struggling with the material (low predicted grade) and those who withdraw because the material was too easy for them (high predicted grade). More unsettling is the distribution of predicted grades for those students who receive F's in M156. From the graph we can see predictions ranging from 0 to slightly above a 3.

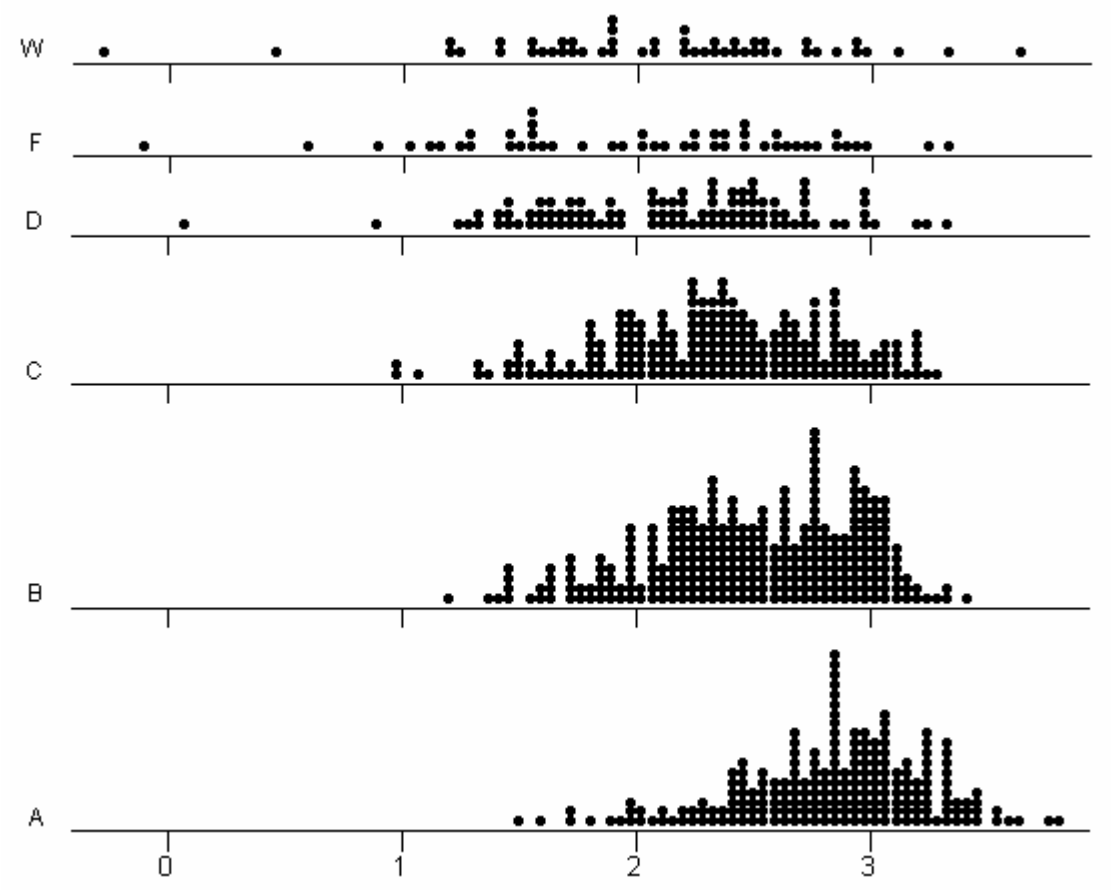

Figure 1 – Predicted grades for students in Math 156 classified by grade received

This extreme variation is indicative of the difficulties in trying to predict a student's success in a single math class based on their math abilities. While math ability (as measured by ACT, high school GPA, and placement tests) is related to success in college mathematics, a large number of intangible factors also lead to success or failure. This simple model does not incorporate such things as study habits, student attitude, emotional adjustments, extra-curricular activities, and various other factors, which are much difficult to quantify and would likely need to be collected post arrival at Truman.

The relatively poor R-sq of the model seems to contradict the graph, which indicates there is valuable information found in the explanatory variables, with higher predicted grades tending to fall into the higher grade categories. The poor model fit results from the fact that students within the data set who have almost identical mathematical skills (as measured by our explanatory variables) manage to achieve all the possible grades. This can be easily seen by reviewing students predicted to receive a B (grade = 3). The majority of students receive an A or B, but some manage to receive a C, D, or even an F. The lower grades most likely due to the extraneous variables mention above.

The phenomenon of achieving all possible grades, for the same ability, leads to another possible statistical analysis of the data. If grades are viewed as categorical classes (rather than numeric values), with A being better than B, and B being better C, etc, then an Ordinal Logistic Regression can be performed. Results of this type of analysis yield a set of recursive equations that can be used to find the probability of falling into each of the ordinal classes. Using the same explanatory variables that were found significant in the stepwise regression results in the following probability equations.

$$\Pi_{WF} = \frac{e^{K}}{1+e^{K}} \qquad K_1 = 5.5591 - 0.09703ACTM - 1.0299SMG - 0.11577PT1$$

$$\Pi_{D} = \frac{e^{K}}{1+e^{K}} - \Pi_{WF} \qquad K_2 = 6.4881 - 0.09703ACTM - 1.0299SMG - 0.11577PT1$$

$$\Pi_C = \frac{e^K}{1+e^K} - \Pi_D \qquad K_3 = 7.7084 - 0.09703ACTM - 1.0299SMG - 0.11577PT1$$

$$\Pi_B = \frac{e^K}{1+e^K} - \Pi_C \qquad K_4 = 9.3552 - 0.09703ACTM - 1.0299SMG - 0.11577PT1$$

$$\Pi_A = 1 - (\Pi_{WF} + \Pi_D + \Pi_C + \Pi_B)$$

These equations are unwieldy for hand calculations but can easily be set up on an EXCEL spreadsheet.

To illustrate the use of the equations, the data from the two students selected before for the regression analysis are shown below. The student with an ACTM of 29, a SMG of 4.0, and a PT1 score of 23 (all indicating high mathematical abilities) would have the following grade probabilities. The individual would have approximately an 86% chance of receiving an A or B in M156, but there would still be a possibility they could receive a C, D, F, or W.

Ability	WF	D	C	B	A	Expected
High	0.01734	0.02543	0.08872	0.30856	0.55992	3.37
Low	0.42569	0.22670	0.21171	0.10648	0.02940	1.08

Table 3 – Predicted grade distribution in M156 using ordinal logistic regression

For comparison a student with a ACTM of 17, a SMG of 3.3, and a PT1 score of 7 (all indicating low mathematical abilities) would have the following grade probabilities and would likely withdraw from the class or receive a D or F in M156.

It is also useful to use these equations to calculate the students "expected" grade. From basic probability theory we know that

$$E(x) = \sum_{i=1}^{n} x_i \cdot p(x_i).$$

Calculating the expected grade for a student with an ACTM of 29, a SMG of 4.0, and a PT1 score of 23 yields an expected grade in M156 of 3.37, while calculating the expected grade for a student with a ACTM of 17, a SMG of 3.3, and a PT1 score of 7 yields an expected grade in M156 of 1.08

It is comforting to note the expected grade for the logistic model is very similar to the expected grade for the regression model. The logistic model, with its probability for each of the various grades, could almost be interpreted at explaining the prediction from the regression model.

2.2 Math 157 (Trigonometry)

The correlation matrix for M157 indicates a much stronger relationship between grade and the various explanatory variables than was seen for students enrolled in M156. All three of the measures of mathematical ability (ACTM, SRG, & PT) have a correlation of approximately 0.45, which is larger than maximum correlation of 0.33 seen in M156. Additionally, the correlation between the 3 explanatory is also stronger indicating the three measures tend to agree in their evaluation of the students mathematical ability.

	Grade	ACTM	Sub 1	Sub 2	Sub 3	PT1	PT2
ACTM	0.457						
Sub 1	0.378	0.753					
Sub 2	0.353	0.676	0.566				
Sub 3	0.400	0.760	0.524	0.490			
PT1	0.447	0.535	0.564	0.483	0.431		
PT2	0.332	0.300	0.279	0.243	0.237	0.591	
SRG	0.453	0.297	0.306	0.196	0.339	0.301	0.112

Table 4 – Correlation matrix of grade and explanatory variables for Math 157

A stepwise regression was conducted to determine which were the strongest predictors of performance in M157. The variables chosen, in order of selection, were SMG, ACTM and then PT1. A regression on these variables yields the following resulting equation.

$$\hat{y} = -4.4055 + 0.09426\,ACTM + 0.06387\,PT1 + 0.9404\,SMG$$

The predicted grades for all students were calculated with the equation and the highest and lowest predictions are listed below. The predicted grade for the high student (4.19) illustrate a slight problem with the model in that the response is not limited to fall between values of 0 and 4. It can best be interpreted as predicting the student has a strong

chance of receiving an A, which they did. The low student is predicted to receive a 0.57 and did withdraw from the class.

Ability	ACTM	PT1	SMG	Predicted	Actual
High	31	30	4.0	4.19	A
Low	24	13	2.0	0.57	W

Table 5 – Predicted versus actual grade in Math 157 utilizing the regression model

All of the model coefficients are highly significant and the model explains 35.6% of the variation in the grade received by the student. Although this leaves a large amount of variation unexplained, the fit for M157 is much better than that found for M156.

The review of the predicted versus actual grade illustrates the improvement in the M157 model over the M156 model. There is an obvious trend in the predicted grades as you progress from A to W. While there is a large amount of overlap in the grade categories, we can see that high predicted grades tend to result in higher grades and low predicted grades result in lower grades.

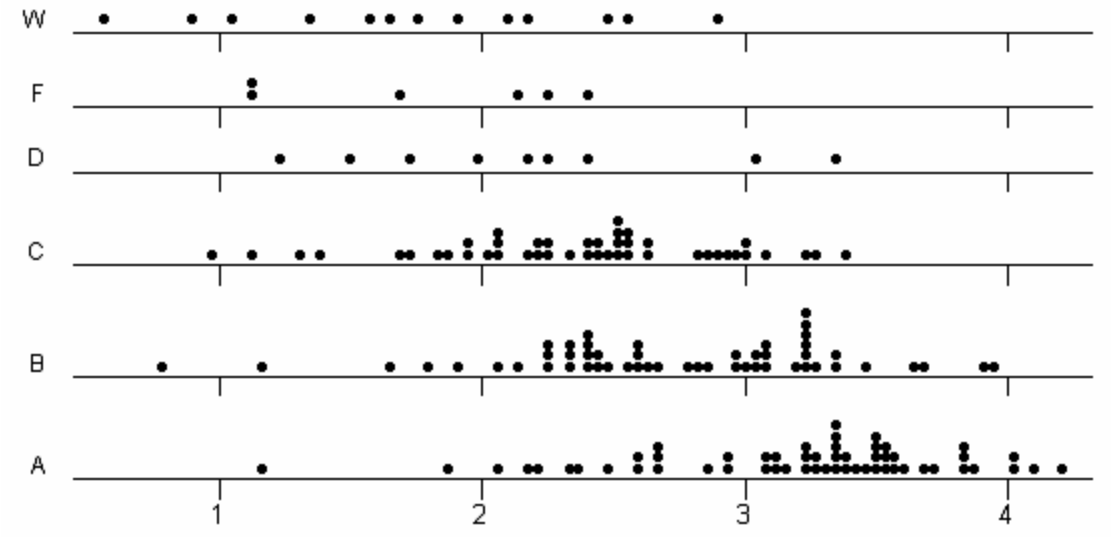

Figure 2 – Predicted grades for students in Math 157 classified by grade received

One point of interest in Figure 2 is the lack of outliers or unusual points for this graph, as no students predicted to do well in M157 received poor grades or withdrew.

An ordinal logistic regression was also calculated for this data using the ACTM, SMG and PT1 as explanatory variables. The recursive equations for the five grade groups can be found below.

$$\Pi_{WF} = \frac{e^K}{1+e^K} \qquad K = 10.680 - 0.19132ACTM - 1.7413SMG - 0.11914PT1$$

$$\Pi_D = \frac{e^K}{1+e^K} - \Pi_{WF} \qquad K = 11.247 - 0.19132ACTM - 1.7413SMG - 0.11914PT1$$

$$\Pi_C = \frac{e^K}{1+e^K} - \Pi_D \qquad K = 12.991 - 0.19132ACTM - 1.7413SMG - 0.11914PT1$$

$$\Pi_B = \frac{e^K}{1+e^K} - \Pi_C \qquad K = 14.639 - 0.19132ACTM - 1.7413SMG - 0.11914PT1$$

$$\Pi_A = 1 - (\Pi_{WF} + \Pi_D + \Pi_C + \Pi_B)$$

Calculating the predicted grade distribution for a student with an ACTM of 31, a SMG of 4.0, and a PT1 score of 30 (high ability student) yields the following grade probabilities. As was seen with the regression model, the student would likely receive an A or B in M157. Calculating the predicted grade distribution for a student with an ACTM of 24, a SMG of 2.0, and a PT1 score of 13 (low ability student) yields the following grade probabilities. The student would likely withdraw from the class or receive a D or F in M157.

Ability	WF	D	C	B	A
High	0.00304	0.00231	0.02455	0.10814	0.86193
Low	0.74208	0.09325	0.13136	0.02671	0.00658

Table 6 – Predicted grade distribution in M157 using ordinal logistic regression

2.3 Math 186 (Elementary Functions)

The correlation matrix for M186 differs from both the M156 and M157 matrices. The strongest linear association with the M186 grade is ACTM and self-reported high school math GPA, which is similar to M156. What is dramatic about the correlations here are that neither PT1 nor PT2, both considered to be important predictors for math placement, are linearly related to the grade received in M186.

	Grade	ACTM	Sub 1	Sub 2	Sub 3	PT1	PT2
ACTM	0.347						
Sub 1	0.280	0.647					
Sub 2	0.241	0.706	0.497				
Sub 3	0.264	0.665	0.459	0.520			
PT1	0.079	0.126	0.127	0.212	0.167		
PT2	0.057	0.032	-0.007	0.080	0.006	0.165	
SRG	0.270	0.098	0.069	0.094	0.116	0.021	0.009

Table 7 – Correlation matrix of grade and explanatory variables for Math 186

In comparing the correlation of the explanatory variables to themselves we see almost no relationship between ACT, placement tests, or self reported GPA. This seems unusual as all of the explanatory variables measure math abilities on some scale. It is important to remember that the individuals in M186 are not randomly selected from a population and the fact that there was conflicting information on an individual may have resulted in them being assigned to this class.

A stepwise regression was conducted to determine which were the strongest predictors of performance in M186. The variables chosen, in order of selection, were ACTM and then SMG. A regression on these variables yields the following resulting equation.

$$\hat{y} = -3.77 + 0.137\,ACTM + 0.694\,SMG$$

Predicting the grades for all M186 students yields the following highest and lowest predicted grades. The high student received an A, which is not unusual for their predicted value but the low student performed much better than predicted. A review of that individuals data did not reveal any information that might have explained their superior performance.

Ability	ACTM	SMG	Predicted	Actual
High	34	4.0	3.66	A
Low	18	2.5	0.43	C

Table 8 – Predicted versus actual grade in Math 186 utilizing the regression model

Once again the model coefficients are highly significant, but the model explains a limited amount of the variation in grades received with only an r-squared of 17.7% of the variation in the grade received by the student.

The review of predicted versus actual grade for this model indicate that students who were predicted to receive good grades tended to receive good grades. Also obvious from the plot is the large amount of overlap that occurs for a specific predicted grade. We can see that a student predicted to get a 2.4 might receive any grade from an A to an F, though they are most likely to get a B. The visual inspection of the model seems to indicate that the M186 model may perform slightly better than the M156 model. If the two unusual high individuals who received W's were removed (as they likely withdrew and transferred to calculus) the graph would show a strong trend.

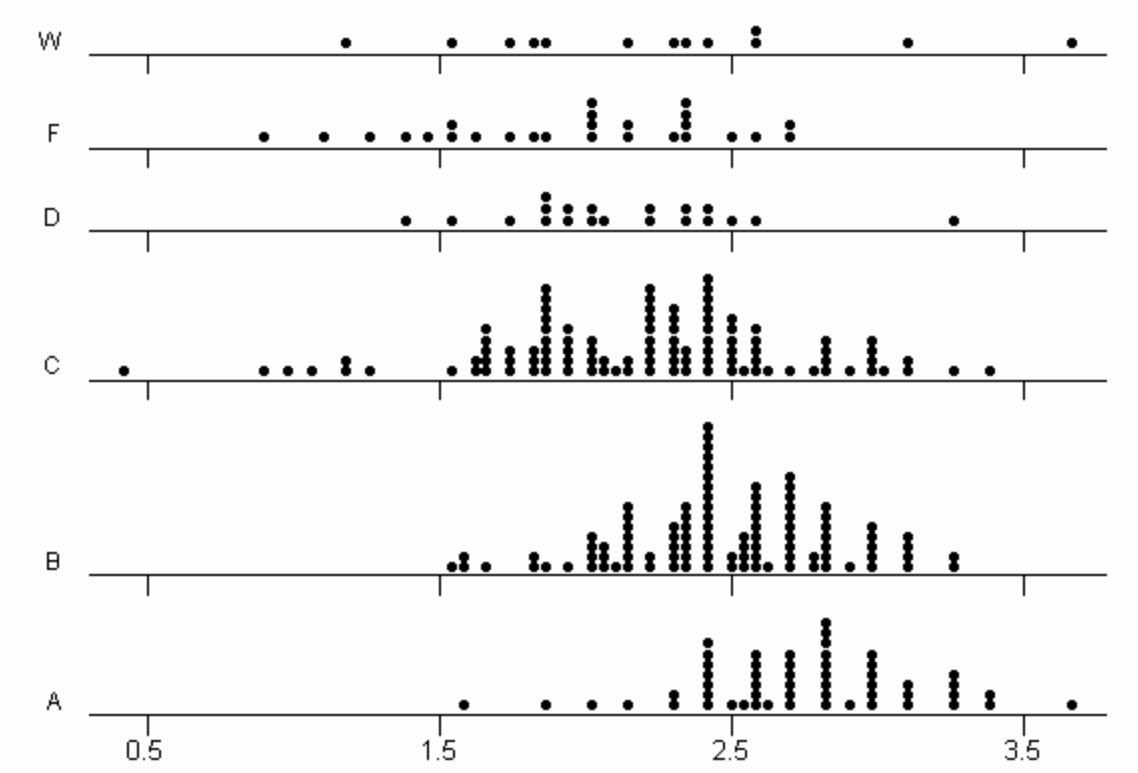

Figure 3 – Predicted grades for students in Math 186 classified by grade received

An ordinal logistic regression was also calculated for the data using the ACTM and SMG as explanatory variables. The recursive equations for the five grade groups can be found below.

$$\Pi_{WF} = \frac{e^K}{1+e^K} \qquad K = 8.809 - 0.25558ACTM - 1.1997SMG$$

$$\Pi_D = \frac{e^K}{1+e^K} - \Pi_{WF} \qquad K = 9.340 - 0.25558ACTM - 1.1997SMG$$

$$\Pi_C = \frac{e^K}{1+e^K} - \Pi_D \qquad K = 11.103 - 0.25558ACTM - 1.1997SMG$$

$$\Pi_B = \frac{e^K}{1+e^K} - \Pi_C \qquad K = 12.865 - 0.25558ACTM - 1.1997SMG$$

$$\Pi_A = 1 - \left(\Pi_{WF} + \Pi_D + \Pi_C + \Pi_B\right)$$

Calculating the predicted grade distribution for a student with an ACTM of 34, and a SMG of 4.0 (both indicating high mathematical abilities) indicates the student would likely receive an A in M186.

Ability	WF	D	C	B	A
High	0.00919	0.00634	0.06870	0.26462	0.65112
Low	0.77017	0.08057	0.12001	0.02408	0.00514

Table 9 – Predicted grade distribution in M186 using ordinal logistic regression

For comparison a student with an ACTM of 18, and a SMG of 2.5 (both indicating lower mathematical abilities) would have the following grade probabilities and would likely withdraw from the class or receive an F in M186. Note the student did have a 12% chance of receiving a C, which they did achieve.

2.4 Math 194 (LAS Calculus)

The correlation matrix for M194 is similar to that found for M156. The self reported GPA has the strongest correlation with grade received while ACTM and PT1 also show linear association as well.

	Grade	ACTM	Sub 1	Sub 2	Sub 3	PT1	PT2
ACTM	0.210						
Sub 1	0.166	0.610					
Sub 2	0.233	0.717	0.426				
Sub 3	0.098	0.699	0.431	0.492			
PT1	0.225	0.399	0.270	0.452	0.296		
PT2	0.160	0.319	0.223	0.343	0.256	.533	
SRG	0.356	0.177	0.106	0.131	0.148	.155	-0.005

Table 10 – Correlation matrix of grade and explanatory variables for Math 194

In comparing the correlation of the explanatory variables to themselves we see a strong relationship between ACTM and the two placement tests.

A stepwise regression was conducted to determine which were the strongest predictors of performance in M194. The variables chosen, in order of selection, were SMG and then PT2. Although ACTM was not strongly significant $(p = 0.16)$ it was kept in the regression as it was used in the other class models. The resulting equation is listed below.

$$\hat{y} = -2.644 + 0.04005ACTM + 1.0546SMG + 0.03708PT2$$

Utilizing the equation and selecting the students with the highest and lowest predicted values yields the following table. We can see that the students preformed close to expectations with the high student receiving and A and the low student receiving a C. The model for M194 has a slightly lower r-square value than the other models with only

14.4% of the variation in the grade received being explained by the 3 explanatory variables.

Ability	ACTM	PT2	SMG	Predicted	Actual
High	33	24	4.0	3.78	A
Low	26	18	2.2	1.38	C

Table 11 – Predicted versus actual grade in Math 194 utilizing the regression model

The predicted versus actual grades does appear to show a weak relationship between predicted grade and earned grade. The relationship is rather weak with only a mild shift in the average predicted grade occurring for each actual grade. There were several individuals who received A's but had low predicted grades (outliers to the left) and further inspection of the data set indicated that the had relatively low self reported grades (2.5-3.5). This would be unexpected as students rarely achieve higher grades in their introductory college mathematics class than their high school marks. This unusual phenomenon may have something to do with nature of LAS Calculus and its concept/application focus.

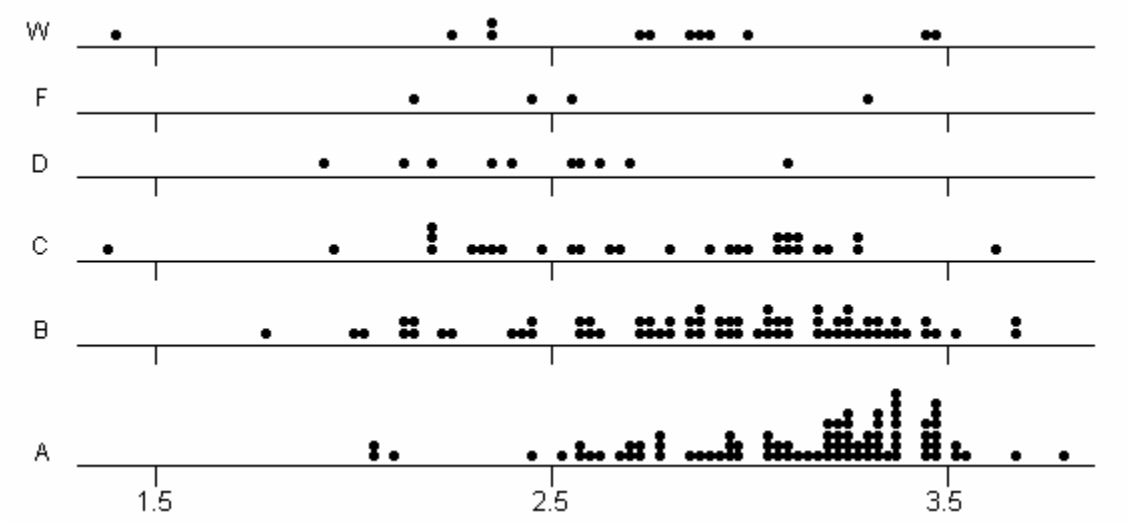

Figure 4 – Predicted grades for students in Math 194 classified by grade received

An ordinal logistic regression was also calculated for the data using the ACTM, SMG, and PT2 as explanatory variables. The recursive equations for the five grade groups can be found below.

$$\Pi_{WF} = \frac{e^K}{1+e^K} \qquad K = 6.847 - 0.08475ACTM - 1.7014SMG - 0.06495PT2$$

$$\Pi_D = \frac{e^K}{1+e^K} - \Pi_{WF} \qquad K = 7.424 - 0.08475ACTM - 1.7014SMG - 0.06495PT2$$

$$\Pi_C = \frac{e^K}{1+e^K} - \Pi_D \qquad K = 8.490 - 0.08475ACTM - 1.7014SMG - 0.06495PT2$$

$$\Pi_B = \frac{e^K}{1+e^K} - \Pi_C \qquad K = 10.087 - 0.08475ACTM - 1.7014SMG - 0.06495PT2$$

$$\Pi_A = 1 - (\Pi_{WF} + \Pi_D + \Pi_C + \Pi_B)$$

Calculating the predicted grade distribution for a student with an ACTM of 33, a SMG of 4.0 and a PT2 of 24 (all indicating high mathematical abilities) indicates the student would likely receive an A in M194.

Ability	WF	D	C	B	A
High	0.01320	0.01007	0.04140	0.18995	0.74535
Low	0.43332	0.14326	0.2214	0.15323	0.04873

Table 12 – Predicted grade distribution in M194 using ordinal logistic regression

For comparison a student with a ACTM of 26, and a SMG of 2.2, and a PT2 of 18 (all indicating lower mathematical abilities) would have the following grade probabilities and would likely withdraw from the class or receive an F in M194.

2.5 Math 198 (Calculus I)

The correlation matrix for the explanatory variables in M198 yields interesting results. As was seen previously both SRG and ACTM are strongly related to grade, but additionally we find that PT2 is also strongly related to performance in M198.

	Grade	ACTM	Sub 1	Sub 2	Sub 3	PT1	PT2
ACTM	0.209						
Sub 1	0.108	0.666					
Sub 2	0.157	0.727	0.501				
Sub 3	0.228	0.692	0.442	0.493			
PT1	0.228	0.311	0.288	0.264	0.278		
PT2	0.295	0.351	0.252	0.266	0.297	0.407	
SRG	0.296	0.089	0.128	0.062	0.063	0.141	0.055

Table 13 – Correlation matrix of grade and explanatory variables for Math 198

A stepwise regression was conducted to determine which were the strongest predictors of performance in M198. The initial variables chosen, in order of selection, were PT2, SMG, and then Sub-3. The fact that PT2 and Sub-3 are significant predictors is interesting as it has been argued that good trigonometric skills are required for success in M198 and both of these variables measure those skills. As the Sub-3 score is missing for about 15% of the data it was determined to remove the sub-score variables and rerun the stepwise regression. With only the cumulative math ACT, placement tests, and self reported GPA available, the best model selected involved PT2, SMG, and ACTM in order of appearance. The elimination of the sub scores for ACT resulted in a very minor decrease in r-squared from to 17.1% to 16.9%. A regression on these variables yields the following resulting equation.

$$\hat{y} = -4.01 + 0.037\,ACTM + 1.01\,SMG + 0.0902\,PT2$$

Utilizing the equation and selecting the students with the highest and lowest predicted values yields the following table. We can see that the students preformed close to expectations with the high student receiving and A and the low student receiving dropping the course.

Ability	ACTM	PT2	SMG	Predicted	Actual
High	35	25	4.0	3.59	A
Low	21	7	3.5	0.94	W

Table 14 – Predicted versus actual grade in Math 198 utilizing the regression model

Once again the coefficients are highly significant, but the model explains a limited amount of the variation in grades received. The r-squared of 16.9% is almost exactly the same as that of the M156 and M186 data sets.

The received versus predicted grade plot indicates a strong trend among the A-B-C range, with actual grade decrease as predicted grade decreased. The distribution of the D-F-W grades are much more spread out with a variety of predicted grades falling into these categories.

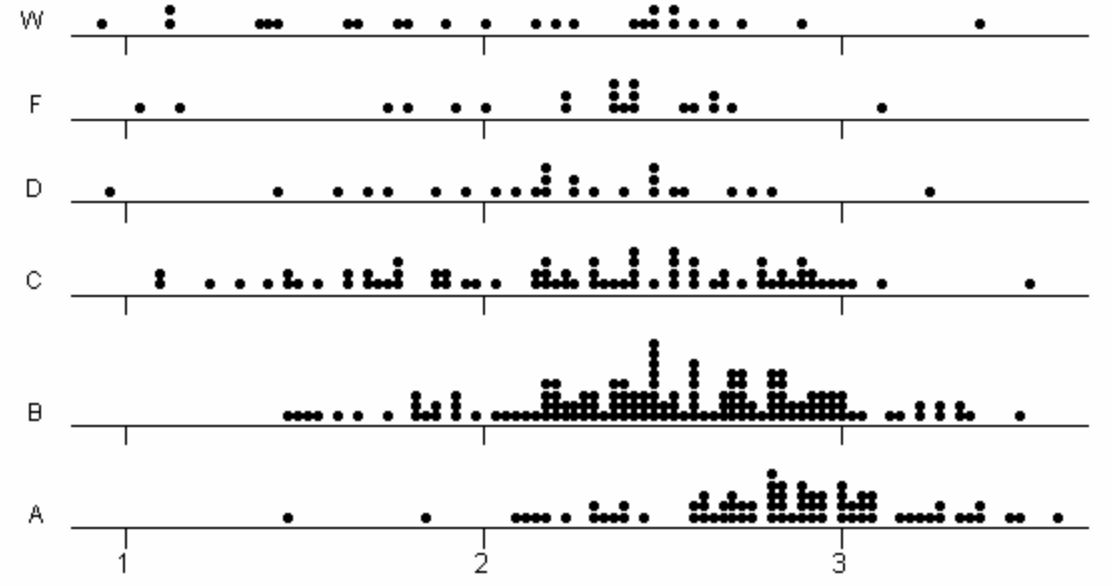

Figure 5 – Predicted grades for students in Math 198 classified by grade received

An ordinal logistic regression was also calculated for the data using the ACTM, PT2, and SMG as explanatory variables. The recursive equations for the five grade groups can be found below.

$$\Pi_{WF} = \frac{e^K}{1+e^K} \qquad K = 8.617 - 0.08248ACTM - 15505SMG - 0.14606PT2$$

$$\Pi_D = \frac{e^K}{1+e^K} - \Pi_{WF} \qquad K = 9.194 - 0.08248ACTM - 15505SMG - 0.14606PT2$$

$$\Pi_C = \frac{e^K}{1+e^K} - \Pi_D \qquad K = 10.347 - 0.08248ACTM - 15505SMG - 0.14606PT2$$

$$\Pi_B = \frac{e^K}{1+e^K} - \Pi_C \qquad K = 12.263 - 0.08248ACTM - 15505SMG - 0.14606PT2$$

$$\Pi_A = 1 - (\Pi_{WF} + \Pi_D + \Pi_C + \Pi_B)$$

Calculating the predicted grade distribution for a student with an ACTM of 35, a SMG of 4.0, and a PT2 of 25 (all indicating high mathematical abilities) indicates the student would likely receive an A in M198.

Ability	WF	D	C	B	A
High	0.01593	0.01207	0.05569	0.29926	0.61703
Low	0.60732	0.12612	0.16370	0.08625	0.016593

Table 15 – Predicted grade distribution in M198 using ordinal logistic regression

For comparison a student with an ACTM of 21, a SMG of 3.5, and a PT2 of 7 (all indicating lower mathematical abilities) would have the following grade probabilities and would likely withdraw or fail and would have a negligible chance of receiving an A.

2.6 Introductory Class Summary

Having developed prediction equations for all introductory classes we are now ready to develop a new placement methodology. The first step is to collect the appropriate data on the student and find a predicted grade for each class. Ideally there would be one class that maximizes the student's expected and that would be their recommended placement. Unfortunately, almost all students will find that their maximum expected grade can be achieved in M156, the most basic math course.

Since starting all students at the lowest level is counter productive to finishing a degree in a timely manner, an alternative approach is suggested. After finding the predicted grade for each class, the student is placed in the highest class that gives them a reasonable chance of success. As an example assume a student with a relatively high mathematical ability was chosen from the data set (ACTM-33, SMG-4, PT1-24, PT2-23).

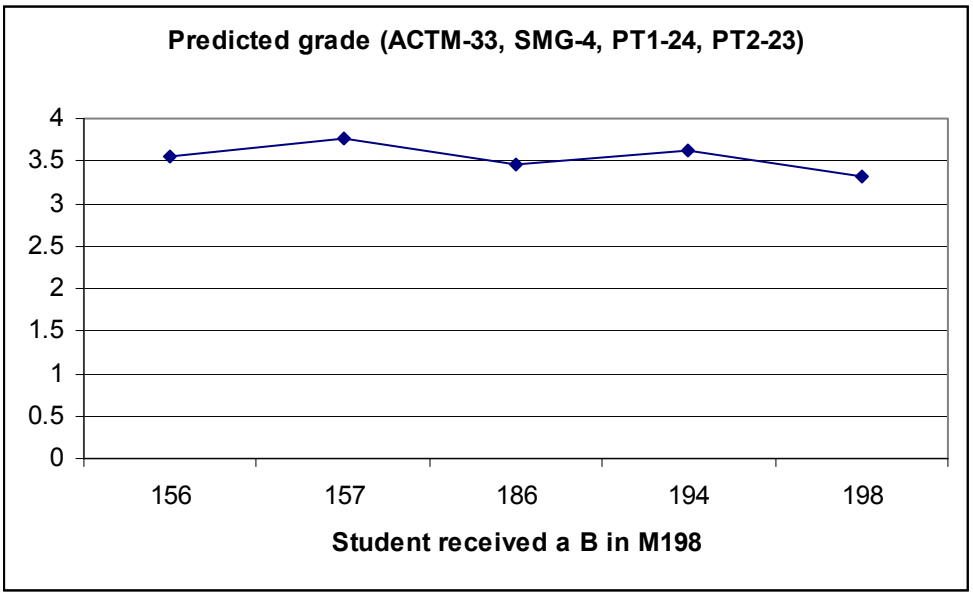

Figure 6 – Predicted grades in the introductory courses for a high ability student

The predicted grade for this student is approximately 3.5 for any of the introductory classes, though a slight decrease in predicted GPA exists for M198. It would seem reasonable to place the student in M198 as they are likely to achieve an A or B and we

would not expect them to do better if placed in a lower class. This particular student in the data set was placed into M198 and did receive a B.

The previous example appears to indicate that this placement methodology is not useful as the student has the same expected grade regardless of the class, but this is strictly a circumstance of the high mathematical abilities of the selected student. If we apply the method to a student with lower mathematical abilities, we can see the value of the method.

Selecting a student from the data set with lower scores (ACTM-24, SMG-3.3, PT1-26, PT2-2) yields the following predicted grades. The predicted grade for this student varies from a 3.0 in M156 to a 0.5 in M198. The large difference in expected grade is a result of the low PT2, which measures trigonometric skills. It would seem reasonable to place this student in M156 or M157 as the expected grades for these two classes are approximately the same. Enrollment in M186 or M194 might be allowed if the student requested this placement, but a student should be strongly dissuaded from enrolling in M198. In this instance the student enrolled into M186 and did receive a C.

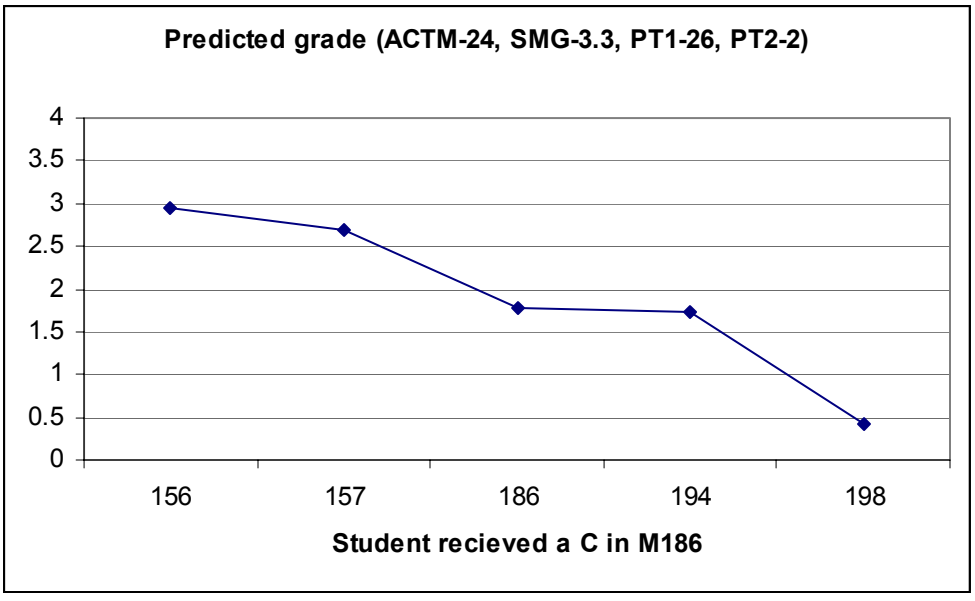

Figure 7 – Predicted grades in the introductory courses for a low ability student

The difficulty with this system is that the placement officer is still required to make some judgment as to what is a "reasonable" decrease in grade, when determining initial placement. Although it is not an ideal system, it does give the placement officer some ability to differentiate those likely to succeed from those likely to fail for a particular math class. This information could even be used proactively to identify students before the first day of class that may need special attention to succeed in their chosen course.

CHAPTER 3 UPPER LEVEL COURSES

3.1 Math 263 (Calculus II)

There are numerous methods for a student to complete Calc I (MATH198) and progress into Calculus II (MATH263) here at Truman. Students can complete Calc I on the Truman campus, transfer credit from another college, utilize dual-credit from high school, or successfully complete the AP exam following a high school calculus course.

For several years, discussion had been occurring within the mathematics department about the "lack of preparation" that some students were exhibiting in Calc II. In theory all students should have the basic skills necessary for success in the course, as proficiency is verified by placement tests or passing marks from the various sources.

Previous to this study no structured attempt had made to explain why some students struggled in the class. A quick review of the many methods of completing Calc I indicates obvious potential sources for the differences in student preparedness. Assuming that the content of Truman Calc I courses were consistent and that Truman had little power to address issues of dual and transfer credit, the focus of the study was directed at investigating the use of high school credit (specifically AP) to fulfill Calc I.

3.2 Advanced Placement (AP)

The Advanced Placement exam consists of two separate exams, the AB exam and the BC exam, with the BC exam being the more advanced of the two. All scores mentioned in this analysis will refer to the AB exam, which represents the material covered during one full year of high school calculus. The AP system issues a numeric grade, rather than the traditional letter grade, with the following definitions:

5 Extremely well qualified
4 Well qualified
3 Qualified
2 Possibly Qualified
1 No recommendation

The AP board recommends scores of 2-5 to receive college credit or advanced placement, while a score of 1 should not receive college credit or advanced placement. (For more information on the exam itself see www.collegeboard.com)

After completing the exam, incoming students have their test scores reported to Truman, where they receive credit for MATH198 if they score a 3 or better.

3.3 Data Collection

The study began by identifying students whose first mathematics enrollment at Truman was Calc II. Focusing on recent years, 330 students were identified as starting in Calc II from 1995-2003 who likely used AP credit for Calc I. As the registrar receives all AP records for the purpose of determining transfer credit, a request was made for the AP exams scores. Unfortunately it was found that the current SIS/Banner systems did not contain the score of the AP exam (1-5) but rather contained only a "T" credit for test. Student scores on the AP exam are recorded but only in paper format found in the advising folder. Normally, the collection of these 330 scores would have required the time consuming task of locating all individual advising folders but fortuitously the Registrar's staff had chosen to store the old AP reports rather than dispose of them after their use.

AP records were found for 244 of the original 330 students. The grades received in Calc II are broken down by AP score in Table 16 below. The table indicates that a score of 4 is slightly more common than either 3 or 5.

	A	B	C	D	F	W	
3	12	17	17	9	3	14	72
4	35	22	20	6	5	19	107
5	36	13	6	1	1	8	65
	83	52	43	16	9	41	244

Table 16 – Number of students receiving various grades for Calc II based on AP score

The raw counts of Table 16 indicate a possible relationship between AP score and class grade and this trend is more evident when the percentages for each grade are calculated for the various AP scores. Table 17 indicates that a students with a low score on the AP exam (3) are much more likely to receive a C, D, or F than their counterparts, with the difference in the number of A's being especially evident.

	A	B	C	D	F	W	
3	16.67	23.61	23.61	12.50	4.17	19.44	100
4	32.71	20.56	18.69	5.61	4.67	17.76	100
5	55.38	20.00	9.23	1.54	1.54	12.31	100
	34.02	21.31	17.62	6.56	3.69	16.80	100

Table 17 – Percentage of students receiving various grades in Calc II by AP exam score

While the lack of A's for the AP-3 group may not seem dramatic, it should be noted that Calc II is a 5-credit class and many of these students would be on scholarship. A 5-credit B (or worse) can make it difficult to achieve the required GPA to remain on scholarship.

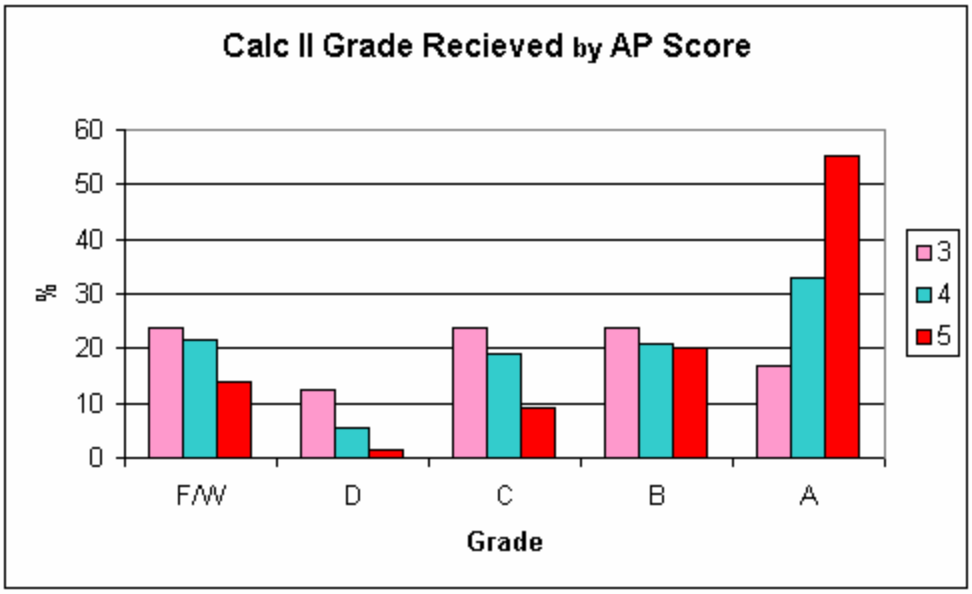

Figure 8 – Percentage of students receiving various grades in Calc II by AP exam score

3.4 Policy Change

These initial results were shared with both the Division Head of Math & Computer Science and the Registrar's Office. In an effort to reward incoming students for advanced work achieved in high school without placing them in a class for which they are unprepared, a revised transfer credit policy was instigated in Fall 2004. The new policy is as follows:

> Students with a score of 4 or 5 on the AB Advanced Placement exam will receive credit for MATH198, while those with a 3 will receive credit for MATH192.

The reduced credit (5 versus 4) seems reasonable as the low score on the AP exam may indicate that the student did not grasp all of the concepts taught in Calc I or possibly that their high school calculus class did not cover the same amount of material that is covered in a 5 credit Truman calculus class.

CHAPTER 4 SUMMARY

Although none of the models considered in this analysis were able to reliably predict a student's actual grade in their first mathematics class, the models were able to give reasonable predictions of success and failure for the various classes. The process of constructing the models and determining significant variables proved to be valuable unto itself as the selection process identifies which of the possible explanatory variables are important for the various classes.

As would be expected the self reported math GPA from high school is one of the most significant predictors of success in college mathematics, with better high school students being better college students. A close second in predictive power is the ACT mathematics score, which proved to be a reliable predictor of success in almost all of the classes. Surprisingly the first placement test was only found useful for the lower level M156 and M157 courses. As the ACTM and this placement exam are fairly redundant there is a possibility that this placement exam could be eliminated from the Truman system. Complete elimination of both placement exams may not be recommend as the second placement exam did prove useful for predicting performance in the calculus classes. The importance of this second exam reaffirms the belief that students should have strong trigonometric skills before they are allowed into the Truman calculus classes.

In general the explanatory variables tended to be correlated with each other. This would be expected as all of the measures considered evaluate mathematical ability at some level. There was very little correlation between the measures and high school GPA but this was most likely due to the large number of entering students who indicate a 4.0 GPA in high school. One interesting discovery amongst the explanatory variables was a lack of correlation for students taking M186. No immediate reason for this phenomenon was found and further investigation is planned.

In summary the models found in this analysis offer a possible alternative to the current system of mathematics placement. The concurrent use of these models with the existing system on a future incoming freshman class would allow for an evaluation of the new system without dramatically changing the current class assignments. Additionally, future work could also include revisions of the data set. Identification of outliers and unusual individuals could have a two-fold benefit. Investigations into factors that resulted in sub-standard or super-standard performances may identify factors Truman can manipulate to help its students succeed. Secondly, the removal/filtering of these unusual observations could improve the predictive power of the models.

APPENDICES

Math 156 – Elementary Functions

Regression Analysis: NGRADE versus ACTM, pt1, srmgpa

```
NGRADE = - 2.22 + 0.0534 ACTM + 0.0702 pt1 + 0.613 srmgpa

895 cases used 5 cases contain missing values

Predictor        Coef     SE Coef          T        P
Constant      -2.2175      0.3492      -6.35    0.000
ACTM          0.05340     0.01317       4.05    0.000
pt1           0.07023     0.01063       6.61    0.000
srmgpa        0.61305     0.07312       8.38    0.000

S = 1.138       R-Sq = 19.1%     R-Sq(adj) = 18.9%

Analysis of Variance

Source            DF          SS          MS         F        P
Regression         3     273.142      91.047     70.24    0.000
Residual Error   891    1154.864       1.296
Total            894    1428.007
```

Ordinal Logistic Regression: NGRADE versus ACTM, srmgpa

```
Link Function:  Logit

Response Information

Variable  Value       Count
NGRADE    0              39
          1              20
          2             102
          3              99
          4              54
          Total         314

  314 cases were used
    1 cases contained missing values

Logistic Regression Table
                                                      Odds        95% CI
Predictor       Coef    SE Coef        Z     P    Ratio    Lower    Upper
Const(1)       8.809      1.358     6.48 0.000
Const(2)       9.340      1.365     6.84 0.000
Const(3)      11.103      1.406     7.89 0.000
Const(4)      12.865      1.448     8.89 0.000
ACTM        -0.25558    0.03921    -6.52 0.000     0.77     0.72     0.84
srmgpa       -1.1997     0.2579    -4.65 0.000     0.30     0.18     0.50

Log-likelihood = -426.078
Test that all slopes are zero: G = 68.738, DF = 2, P-Value = 0.000

Goodness-of-Fit Tests

Method      Chi-Square    DF      P
Pearson        409.930   358  0.030
Deviance       333.755   358  0.817
```

Measures of Association:
(Between the Response Variable and Predicted Probabilities)

Pairs	Number	Percent	Summary Measures	
Concordant	25489	69.3%	Somers' D	0.41
Discordant	10505	28.6%	Goodman-Kruskal Gamma	0.42
Ties	783	2.1%	Kendall's Tau-a	0.30
Total	36777	100.0%		

Math 157 – Trigonometry

Regression Analysis: NGRADE versus ACTM, srmgpa, pt1

```
NGRADE = - 4.41 + 0.0943 ACTM + 0.940 srmgpa + 0.0639 pt1

181 cases used 2 cases contain missing values

Predictor        Coef     SE Coef          T        P
Constant      -4.4055      0.7638      -5.77    0.000
ACTM          0.09426     0.02902       3.25    0.001
srmgpa         0.9404      0.1863       5.05    0.000
pt1           0.06387     0.02075       3.08    0.002

S = 1.019       R-Sq = 35.6%     R-Sq(adj) = 34.5%

Analysis of Variance

Source            DF          SS          MS         F        P
Regression         3     101.646      33.882     32.64    0.000
Residual Error   177     183.724       1.038
Total            180     285.370
```

Ordinal Logistic Regression: NGRADE versus ACTM, srmgpa, pt1

```
Link Function:  Logit

Response Information

Variable  Value       Count
NGRADE    0              19
          1               9
          2              45
          3              52
          4              56
          Total         181

  181 cases were used
    2 cases contained missing values

Logistic Regression Table
                                                        Odds         95% CI
Predictor       Coef    SE Coef        Z     P    Ratio    Lower    Upper
Const(1)      10.680      1.607     6.65 0.000
Const(2)      11.247      1.622     6.93 0.000
Const(3)      12.991      1.693     7.68 0.000
Const(4)      14.639      1.762     8.31 0.000
ACTM        -0.19132    0.05558    -3.44 0.001     0.83     0.74     0.92
srmgpa       -1.7413     0.3634    -4.79 0.000     0.18     0.09     0.36
pt1         -0.11914    0.03893    -3.06 0.002     0.89     0.82     0.96

Log-likelihood = -222.574
Test that all slopes are zero: G = 80.901, DF = 3, P-Value = 0.000

Goodness-of-Fit Tests

Method      Chi-Square    DF      P
Pearson        599.686   609  0.598
Deviance       390.550   609  1.000
```

Measures of Association:
(Between the Response Variable and Predicted Probabilities)

Pairs	Number	Percent	Summary Measures	
Concordant	9474	77.5%	Somers' D	0.56
Discordant	2680	21.9%	Goodman-Kruskal Gamma	0.56
Ties	73	0.6%	Kendall's Tau-a	0.42
Total	12227	100.0%		

Math 186 – Elementary Functions

Regression Analysis: NGRADE versus ACTM, srmgpa

```
NGRADE = - 3.77 + 0.137 ACTM + 0.694 srmgpa

314 cases used 1 cases contain missing values

Predictor        Coef     SE Coef          T        P
Constant      -3.7715      0.7533      -5.01    0.000
ACTM          0.13685     0.02187       6.26    0.000
srmgpa         0.6939      0.1494       4.64    0.000

S = 1.095       R-Sq = 17.7%     R-Sq(adj) = 17.2%

Analysis of Variance

Source            DF          SS          MS         F        P
Regression         2      80.406      40.203     33.54    0.000
Residual Error   311     372.756       1.199
Total            313     453.162

Source       DF      Seq SS
ACTM          1      54.559
srmgpa        1      25.847
```

Ordinal Logistic Regression: NGRADE versus ACTM, srmgpa

```
Link Function:  Logit

Response Information

Variable  Value       Count
NGRADE    0              39
          1              20
          2             102
          3              99
          4              54
          Total         314

  314 cases were used
    1 cases contained missing values

Logistic Regression Table
                                                    Odds        95% CI
Predictor       Coef    SE Coef        Z     P    Ratio    Lower    Upper
Const(1)       8.809      1.358     6.48 0.000
Const(2)       9.340      1.365     6.84 0.000
Const(3)      11.103      1.406     7.89 0.000
Const(4)      12.865      1.448     8.89 0.000
ACTM        -0.25558    0.03921    -6.52 0.000     0.77     0.72     0.84
srmgpa       -1.1997     0.2579    -4.65 0.000     0.30     0.18     0.50

Log-likelihood = -426.078
Test that all slopes are zero: G = 68.738, DF = 2, P-Value = 0.000
```

Goodness-of-Fit Tests

Method	Chi-Square	DF	P
Pearson	409.930	358	0.030
Deviance	333.755	358	0.817

Measures of Association:
(Between the Response Variable and Predicted Probabilities)

Pairs	Number	Percent	Summary Measures	
Concordant	25489	69.3%	Somers' D	0.41
Discordant	10505	28.6%	Goodman-Kruskal Gamma	0.42
Ties	783	2.1%	Kendall's Tau-a	0.30
Total	36777	100.0%		

Math 194 – LAS Calculus

Regression Analysis: NGRADE versus ACTM, srmgpa, pt2

```
NGRADE = - 2.64 + 0.0400 ACTM + 1.05 srmgpa + 0.0371 pt2

206 cases used 8 cases contain missing values

Predictor        Coef     SE Coef          T        P
Constant       -2.644       1.022      -2.59    0.010
ACTM          0.04005     0.02854       1.40    0.162
srmgpa         1.0546      0.2179       4.84    0.000
pt2           0.03708     0.01957       1.89    0.060

S = 1.118       R-Sq = 14.4%     R-Sq(adj) = 13.1%

Analysis of Variance

Source            DF          SS          MS         F        P
Regression         3      42.353      14.118     11.30    0.000
Residual Error   202     252.404       1.250
Total            205     294.757
```

Ordinal Logistic Regression: NGRADE versus ACTM, srmgpa, pt2

```
Link Function:  Logit

Response Information

Variable  Value       Count
NGRADE    0              16
          1              10
          2              30
          3              68
          4              82
          Total         206

  206 cases were used
    8 cases contained missing values

Logistic Regression Table
                                                        Odds        95% CI
Predictor       Coef    SE Coef        Z     P    Ratio    Lower    Upper
Const(1)       6.847      1.756     3.90 0.000
Const(2)       7.424      1.759     4.22 0.000
Const(3)       8.490      1.780     4.77 0.000
Const(4)      10.087      1.822     5.54 0.000
ACTM        -0.08475    0.04786    -1.77 0.077     0.92     0.84     1.01
srmgpa       -1.7014     0.3719    -4.57 0.000     0.18     0.09     0.38
pt2         -0.06495    0.03256    -1.99 0.046     0.94     0.88     1.00

Log-likelihood = -262.773
Test that all slopes are zero: G = 34.138, DF = 3, P-Value = 0.000
```

Goodness-of-Fit Tests

Method	Chi-Square	DF	P
Pearson	717.973	689	0.215
Deviance	475.272	689	1.000

Measures of Association:
(Between the Response Variable and Predicted Probabilities)

Pairs	Number	Percent	Summary Measures	
Concordant	10087	67.6%	Somers' D	0.36
Discordant	4688	31.4%	Goodman-Kruskal Gamma	0.37
Ties	141	0.9%	Kendall's Tau-a	0.26
Total	14916	100.0%		

Math 198 – Calculus I

Regression Analysis: NGRADE versus ACTM, srmgpa, pt2

```
NGRADE = - 4.01 + 0.0370 ACTM + 1.01 srmgpa + 0.0902 pt2

352 cases used 5 cases contain missing values

Predictor        Coef     SE Coef          T        P
Constant      -4.0145      0.8945      -4.49    0.000
ACTM          0.03705     0.02285       1.62    0.106
srmgpa         1.0132      0.1847       5.49    0.000
pt2           0.09024     0.01881       4.80    0.000

S = 1.166       R-Sq = 16.9%     R-Sq(adj) = 16.1%

Analysis of Variance

Source            DF          SS          MS         F        P
Regression         3      95.980      31.993     23.53    0.000
Residual Error   348     473.199       1.360
Total            351     569.179
```

Ordinal Logistic Regression: NGRADE versus ACTM, srmgpa, pt2

```
Link Function:  Logit

Response Information

Variable  Value       Count
NGRADE    0              47
          1              26
          2              74
          3             131
          4              74
          Total         352

  352 cases were used
    5 cases contained missing values

Logistic Regression Table
                                                    Odds        95% CI
Predictor       Coef    SE Coef        Z     P    Ratio    Lower    Upper
Const(1)       8.617      1.462     5.89 0.000
Const(2)       9.194      1.469     6.26 0.000
Const(3)      10.347      1.491     6.94 0.000
Const(4)      12.263      1.528     8.02 0.000
ACTM        -0.08248    0.03606    -2.29 0.022     0.92     0.86     0.99
srmgpa       -1.5505     0.2970    -5.22 0.000     0.21     0.12     0.38
pt2         -0.14606    0.03040    -4.80 0.000     0.86     0.81     0.92

Log-likelihood = -486.737
Test that all slopes are zero: G = 71.883, DF = 3, P-Value = 0.000
```

Goodness-of-Fit Tests

Method	Chi-Square	DF	P
Pearson	998.614	961	0.194
Deviance	753.294	961	1.000

Measures of Association:
(Between the Response Variable and Predicted Probabilities)

Pairs	Number	Percent	Summary Measures	
Concordant	31844	68.6%	Somers' D	0.38
Discordant	14249	30.7%	Goodman-Kruskal Gamma	0.38
Ties	360	0.8%	Kendall's Tau-a	0.28
Total	46453	100.0%		