MEASUREMENT OF ACCUARACY OF THE ADVANCED PLACEMENT EXAM FOR CALCULUS II AND THE INFLUENTIAL FACTORS TO PRECALCULUS

By

Carolyn Swango

A SENIOR RESEARCH PAPER PRESENTED TO THE DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE OF STETSON UNIVERSITY IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF BACHELOR OF SCIENCE

STETSON UNIVERSITY

2005

TABLE OF CONTENTS

LIST OF FIGURES

FIGURES

ABSTRACT

MEASUREMENT OF ACCUARACY OF THE ADVANCED PLACEMENT EXAM FOR CALCULUS II AND THE INFLUENTIAL FACTORS TO PRECALCULUS

By

Carolyn Swango

May 2005

Advisor: Dr. Will Miles
Department: Mathematics and Computer Science

Standardized test scores are becoming an ever-increasing concern for many educators. One of the concerns that have been raised is whether these standardized test scores are as effective as a regular course. The Advanced Placement Calculus Exam is an exam that is used at the high school level to assess the progress of a student's ability in Calculus. This research study will compare the AP Calculus Exam to that of a Calculus I course. Hypothesis test are constructed that compare AP Calculus Exam scores to that of Calculus I grades. In addition to an analysis of the AP Calculus Exam, this research study will also include an analysis of Precalculus students. By analyzing different demographics of Precalculus students, determining which of these variables may be more influential towards a Precalculus grade. By building different regression models and then determining an R^2 statistic can assist in determining which variables may be more essential for a Precalculus grade.

CHAPTER 1
INTRODUCTION: AP CALCULUS STUDY

The goal of a standardized test is to "measure students by a single standard" [B1]. Many students who enter college come from different educational backgrounds. A standardized test is a method of taking students from across the nation and measuring their abilities by means of a common tool. Through the years much controversy has arisen as a result of standardized tests. Many educators have voiced concern on whether these tests are accurate measurements of a student's proficiency.

To begin with, the Advanced Placement Exams were designed to allow students "to take college-level courses and exams and to earn college credit or placement while still in high school" [B2]. Since, the AP Calculus Exam is a standardized exam that focuses on topics in Calculus; many college professors have developed concern that these exam scores may not be an accurate projector for a student's level of preparedness for his or her next Calculus class. In other words, the scores that are received on this exam are possibly inflated. This idea proposes an interesting study of the AP Calculus Exam. Is the AP Calculus Exam an adequate measure of preparedness for Calculus II compared to that of a Calculus I course at a university? This research study investigates this question.

To begin with, it is important to be familiar with the Advanced Placement Program. This program has been used throughout the United States for many years. In 1955, the beginnings of the current Advanced Placement Program were born. The Ford Foundation Fund developed two different studies: the General Education in School and College and the Kenyon Plan. It was from these two studies that the Advanced Placement Program was born. In that same year, the Ford Foundation Fund for the Advancement of Education organization established "eleven initial subject offerings: American history, biology, chemistry, English, French, German, Latin IV (fourth-year Vergil), Latin V (prose, comedy, lyric), mathematics, physics, and Spanish" [B2]. It was not until 1961 that an increase number of colleges and universities began to honor the AP Exam grades of a three or better [B2].

In this study, the exam that will be discussed and analyzed will be that of the AP mathematics, or the Advanced Placement Calculus Exam. Through the course of the years AP mathematics has been revised and changed. In 1969, the AP mathematics exam was split into the Calculus AB and Calculus BC. The Calculus AB exam focuses on topics covered in Calculus I, and the Calculus BC exam focuses on topics covered in Calculus II. The content of these exams has remained fairly constant through the years. However, due to the increase in technology and popularity, the exam has under gone some revisions.

In 1995, with the popularity of the graphing calculator, that the AP Calculus Exam was revised. The writers of the exam added a section formatted for the use of the graphing calculator. In this section, the students had to be familiar with the technology in order to better perform on the exam. With the addition of the calculator, a different form of education had been introduced. As a result, a new concern over the validity of a student's knowledge in the subject of Calculus has arisen.

In addition to the history of the development of the AP Calculus exam, it is also important to understand the format of the exam. The AP Calculus curriculum is designed "a full high school academic year of work that is comparable to calculus courses in colleges and universities" [B2]. The topics that are tested on the AP Calculus AB exam include:

1. Functions, Graphs, and Limits
    a. Analysis of graphs
    b. Limits of Functions (including one-sided limits)
    c. Asymptotic and unbounded behavior
    d. Continuity as a property of functions
2. Derivatives

   a. Concept of the derivative
   b. Derivative at a point
   c. Derivative as a function
   d. Second derivatives
   e. Applications of derivatives
   f. Computation of derivatives
3. Integrals
   a. Interpretations and properties of definite integrals
   b. Applications of integrals
   c. Fundamental Theorem of Calculus
   d. Techniques of antidifferentiation
   e. Applications of antidifferentiation
   f. Numerical approximations of definite integrals

[B2]

These are common topics that are covered in the first year of a college Calculus class. Thus, the AP Calculus AB Exam, if designed appropriately, should be a valid indicator of the mastery of topics covered in Calculus I.

The AP exam is given scores of 1, 2, 3, 4, or 5. Students who receive a 5 are considered extremely well qualified. Those who receive a 4 are considered well qualified, a score of 3 are considered qualified, a score of 2 is considered possibly qualified, and finally students who receive a score of 1 are considered no recommendation [B2]. In short, students who receive a 3, 4, or 5 pass the exam. Stetson University only grants credit to those students who receive a score of a 4 or 5 on the AP Calculus Exam, whereas the University of Central Florida grants credit to a 3 or above. However in order to combine data from these two universities, this study includes only those students who have received a score of a 4 or a 5 on the AP Calculus Exam.

The exam scores are computed as follows. "Scores on the essay [free response] and problem-solving questions [these] are combined with the results of the computer-scored multiple-choice questions, and the total raw scores are converted to AP's 5-point scale" [B2]. The multiple choice section score is determined by "how many answers the student got wrong, then deduct[ing] a fraction of that number from the number of right answers" [B2]. Also, the scoring of the free-response section of the AP Calculus exam is graded by AP readers. An important aspect of the scoring of this is exam is the process of checks and balances that is applied throughout these sections [B2]. This process shows the validity of a student's score. The goal of this study is to determine if the Advanced Placement Calculus Exam is a reasonable indicator of a student's preparedness for Calculus II. In other words, is the AP Calculus course comparable to the actual Calculus I course taught at a university?

CHAPTER 2
DATA COLLECTION

In this portion of the study, the main focus is on the Advanced Placement Calculus Exam and determining if it is a reasonable indicator of a student's preparedness for Calculus II. The data that is used in this study is from both Stetson University and the University of Central Florida.

The first data set collected was from Stetson University. The Stetson University office of Institutional Research provided a great amount of assistance in gathering the data used in this part of the study. There have been numerous meetings with both John Tichenor and Patti Sanders. Their assistance aided in determining the format of the data that is used. The data that was collected included records of students who took Calculus II. Each data record includes the following categories: SAT Math and Verbal Scores, High School GPA, major at resident university, cumulative college GPA, gender, race, state of residence, age, and the grades of Calculus courses that have been indicated that credit has been granted.

During this process, the institutional research office had difficulties gathering the data in an organized fashion. Each student's information was stored on multiple lines, thereby making it difficult to read. A computer program was needed that would reformat the data into a single line for each student. The final data set that was provided by the office of Institutional Research included a sample size of 337 students.

In addition to the data that was provided by Stetson University, the University of Central Florida also provided a substantial amount of data. There were many email exchanges with Dr. Margaret Borden and Patti Ramsey of the University of Central Florida regarding what data was needed for this study. The University of Central Florida was able to provide two different sets of data. One data set was of those students who had taken and received credit for the AP Calculus

Exam and then took Calculus II. The other data set was of those students who took Calculus I and Calculus II at the University of Central Florida. The AP Calculus Exam data set had a sample size of 1677 students. The Calculus I university students had a sample size of 224 students. The combined sample size of the University of Central Florida is of 1901 students. The final sample size is 2238 students.

Once this data was received, sorting and parsing were needed in order to extract those students who directly related to this study. The University of Central Florida offers different Calculus courses; thereby it became important to identify which course was similar to that of Stetson University's Calculus course. This was figured by the topics that were covered in the class, along with the textbook that was used for the class. Thus, the Calculus courses that are analyzed use the same textbook and cover the same concepts. Once the related courses were identified, the corresponding student records were selected to be included in the study.

After the data collected from Stetson University and the University of Central Florida were sorted, it became important to combine the data to have one sample set that will be analyzed. The data was then reviewed and divided into two categories: those students who took the Advanced Placement Calculus Exam and those students who received credit for Calculus I from a university course.

As a result, the first group consists of students who have received credit for Calculus I by receiving a score of a 4 or 5 on the Advanced Placement Exam. The second group consists of students who received credit for taking Calculus I from a university with a grade of either an A or a B. Since, Stetson University grants credit to only those students who scored a 4 or 5 on the AP Calculus Exam, those University of Central Florida students who received a 3 were removed from the records under consideration. Similarly, only students who earned an A or B were concluded in the second group of records. The assumption that is to be made is that a 4 and a 5 compare directly to a B and an A.

In summary, the data that is used in this study is divided into two groups: first, students who received a 4 or a 5 on the Advanced Placement Calculus Exam and contributed to complete Calculus II and second, students who took Calculus I, received either an A or a B and continued to complete Calculus I. From here forward in the study the first group will be denoted by AP and the second group will be denoted by U.

CHAPTER 3
DATA ANALYSIS: DISTRIBUTIONS

The conclusions regarding the Advanced Placement Exam will be based on a statistical analysis comparing Group AP to Group U. To begin the comparison, a histogram was constructed to show the Calculus II grades distribution for each group.

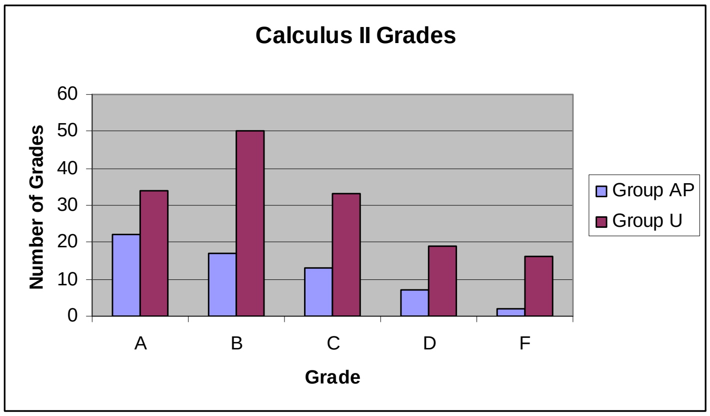


Figure [1]: Calculus II Grades

Since the distributions of the two groups do not appear to be of the same character, it is appropriate to conduct a more detailed analysis.

In order to determine if the given data fits the normal distribution, it became necessary to use the Chi-Squared Goodness of Fit test. The Chi-Squared Goodness of Fit test is a method to test how well observed data fits expected data [A5]. In order to analyze this observed data the expected data must be computed. This was first completed by dividing the grades into five groups, or bins. Grade A represented values from $(3.5, \infty)$; B is $(2.5, 3.5)$; C is $(1.5, 2.5)$; D is $(0.5, 1.5)$ and F is $(-\infty, 0.5)$. The next task was to compute the expected value for each bin. The method to compute the expected values began by first evaluating the normal distribution probability density function for each different bin. "A variable X has a normal distribution if its p.d.f. is defined by $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]}, -\infty < x < \infty$," [A2]. When $\int \frac{1}{\sigma\sqrt{2\pi}} e^{\left[\frac{(x-\mu)^2}{2\sigma^2}\right]} dx$ is bounded by the appropriate bin and then evaluated the result is the probability that occurs for each bin indicated. The expected number is the probability multiplied by the sample size, *n*. In the equation, the values of μ and σ are the mean of the sample size and the standard deviation, respectively" [A2].

For Group AP:

$$\mu = 2.819672 \qquad \sigma = 1.14758$$

$$\begin{aligned}
\int_{-\infty}^{0.5} \frac{1}{(1.14758)\sqrt{2\pi}} e^{\left[\frac{(x-2.819672)^2}{2(1.14758)^2}\right]} dx &= 0.0216213 \\
\int_{0.5}^{1.5} \frac{1}{(1.14758)\sqrt{2\pi}} e^{\left[\frac{(x-2.819672)^2}{2(1.14758)^2}\right]} dx &= 0.103459 \\
\int_{1.5}^{2.5} \frac{1}{(1.14758)\sqrt{2\pi}} e^{\left[\frac{(x-2.819672)^2}{2(1.14758)^2}\right]} dx &= 0.265211 \\
\int_{2.5}^{3.5} \frac{1}{(1.14758)\sqrt{2\pi}} e^{\left[\frac{(x-2.819672)^2}{2(1.14758)^2}\right]} dx &= 0.333064 \\
\int_{3.5}^{\infty} \frac{1}{(1.14758)\sqrt{2\pi}} e^{\left[\frac{(x-2.819672)^2}{2(1.14758)^2}\right]} dx &= 0.276645
\end{aligned}$$

[2]

Thus the expected values are computed as follows:

Normal Distribution		Group AP				
		observed	expected	difference	difference squared	chi value
A	3.5 and greater	22	16.87535	5.124655	26.26209	1.55624
B	3.5-2.5	17	20.3169	-3.3169	11.00185	0.541512
C	2.5-1.5	13	16.17787	-3.17787	10.09886	0.624239
D	1.5-.5	7	6.310999	0.689001	0.474722	0.075221
F	.5 and smaller	2	1.318899	0.681101	0.463898	0.351731
Total:	5 bins	61		sum of chi values:		3.148944

[3]

In order to determine if these observed values are appropriate for the normal distribution the chi-squared statistic must be determined. This is done by finding the difference between the observed and expected values then squaring their difference the chi-values are computed by the formula $\sum_{n=1}^{k}\frac{(\text{observed}_n - \text{expected}_n)^2}{\text{expected}}$ [A5]. This sum is called the chi-squared statistic and it is analyzed. The same method is applied to Group U.

For Group U:

$\mu = 2.440789$
$\sigma = 1.259315$

$$\int_{-\infty}^{0.5}\frac{1}{(1.259315)\sqrt{2\pi}}e^{\left[\frac{(x-2.440789)^2}{2(1.259315)^2}\right]}dx = 0.0616406$$

$$\int_{0.5}^{1.5}\frac{1}{(1.259315)\sqrt{2\pi}}e^{\left[\frac{(x-2.440789)^2}{2(1.259315)^2}\right]}dx = 0.165872$$

$$\int_{1.5}^{2.5}\frac{1}{(1.259315)\sqrt{2\pi}}e^{\left[\frac{(x-2.440789)^2}{2(1.259315)^2}\right]}dx = 0.291238$$

$$\int_{2.5}^{3.5}\frac{1}{(1.259315)\sqrt{2\pi}}e^{\left[\frac{(x-2.440789)^2}{2(1.259315)^2}\right]}dx = 0.281104$$

$$\int_{3.5}^{\infty}\frac{1}{(1.259315)\sqrt{2\pi}}e^{\left[\frac{(x-2.440789)^2}{2(1.259315)^2}\right]}dx = 0.200146$$

[4]

Normal Distribution		Group U				
		observed	expected	difference	diff sq	chi value
A	3.5 and greater	34	30.42219	3.577808	12.80071	0.420769
B	3.5-2.5	50	42.72781	7.272192	52.88478	1.237713
C	2.5-1.5	33	44.26818	-11.2682	126.9718	2.868241
D	1.5-.5	19	25.21254	-6.21254	38.5957	1.530814
F	0.5 and smaller	16	9.369371	6.630629	43.96524	4.692443
total:	5 bins	152		sum of chi values:		10.74998

[5]

The chi-squared value, for Group AP is 3.148944, with 4 degrees of freedom and with a significance of 0.05, this corresponds to a value of 9.488. This means that there is a 5% chance that it would be appropriate to expect more than 5% of the time a value of chi-squared that is as large or larger than 3.148944. This can lead to an indication that the observed values agree with the expected values. In Group U, the chi-squared statistic is 10.74998. Using a significance level of 0.05, with 4 degrees of freedom, this corresponds to the value of 9.488. It is appropriate to say that Group U observed values would not agree with the expected values [A1]. However, another test may be needed to produce results.

An additional way to determine if these two proportions come from similar or different distributions is by the *U*-test. The *U*-test is a nonparametric test that allows two proportions to be studied without assuming any particular distribution. The result of this test will be able to provide information in determining if the distributions of Group AP and Group U are the same or different.

In order to perform a hypothesis test using the *U*-test, it is essential to establish some beginning concepts. The data from the two groups must be combined into one sample set and then ranked. The data was compiled into a spreadsheet in order to assign a rank. In this particular example, there are only five grades, A, B, C, D, and F. The ranks are determined by listing all of the data from smallest to largest, and then assign each grade a number from 1 to 213. Since, there are multiple records of the same grade it is important to determine the ranks based off of a tie. The resulting ranks are as follows:

Rank: F D C B A

$$\frac{\sum_{i=1}^{18} i}{18} = 9.5 \qquad \frac{\sum_{i=19}^{44} i}{26} = 31.5 \qquad \frac{\sum_{i=45}^{90} i}{46} = 67.5 \qquad \frac{\sum_{i=91}^{157} i}{67} = 124 \qquad \frac{\sum_{i=158}^{213} i}{56} = 185.5 \qquad [6]$$

After ranking has been determined for each grade, the next step is to analyze each of the groups separately. This is done by summing all of the ranks for each of the values in the group, this is *W*. After this sum is determined, the *U* values can be found. They are determined by the following formula: $U = W - \frac{n(n+1)}{2}$.

Group AP:

$$n_{ap} = 61$$
$$W_{ap} = 7306$$
$$U_{ap} = W_{ap} - \frac{n_{ap}(n_{ap}+1)}{2} = 7306 - \frac{61(61+1)}{2} = 5415$$

Group U:

$$n_u = 152$$
$$W_u = 15485$$
$$U_u = W_u - \frac{n_u(n_u+1)}{2} = 15485 - \frac{152(152+1)}{2} = 3857$$

In order to continue with the *U*-test, it is necessary to state that when the two sample sizes are larger than eight, "it is considered reasonable to assume that U_{ap} and U_u are values of random variables having approximately normal distribution. To perform the *U*-test based on this assumption the following is needed:

Theorem: Under the assumptions required by the *U*-test, U_{ap} and U_u are values of random variables having the mean $\mu = \frac{n_1 n_2}{2}$ and the variance $\sigma^2 = \frac{n_1 n_2 (n_1 + n_2 + 1)}{12}$," [A4].

Applying this theorem for the given data:

$$\mu = \frac{(61)(152)}{2} = 4636$$
$$\sigma^2 = \frac{(61)(152)(61+152+1)}{12} = 165350.6667\,.$$
$$\sigma = 406.63332$$

Since, $U_{ap} + U_u$ is always equal to $n_{ap}n_u$, the mean and the variance of the random variable corresponding to U_u are equal to those of the random variable corresponding to U_{ap} [A4].

In determining whether or not to reject or not reject a null hypothesis the critical region is determined by the value of α. It has been stated that value of $\alpha = 0.01$ is considered highly significant and a value of $\alpha = 0.05$ is considered significant. The following hypothesis test is a two-tailed test and has a critical region of $\frac{\alpha}{2}$ [A1].

The *U*-test:

$$H_0 : \mu_{ap} = \mu_u$$
$$H_a : \mu_{ap} \neq \mu_u$$
$$\alpha = .10$$
$$\frac{\alpha}{2} = .05$$

[7]
[A4]

Reject the null hypothesis if $z \neq 1.645$, where $z = \frac{U_{ap} - \mu}{\sigma}$.

Thus, $z = \frac{U_{ap} - \mu}{\sigma} = \frac{5415 - 4636}{406.63332} = 1.915730804$.

Since $z = 1.915730804$ and does not equal 1.645, then it is appropriate to reject the null hypothesis. By rejecting the null hypothesis, the two populations do not have the same mean, which implies that Group AP and Group U come from two different distributions.

CHAPTER 4
DATA ANALYSIS: GRADES

In Group AP the sample size is sixty-one students, or $n_{ap} = 61$. The number of students who received each grade is as follows:

Grade	Quantity
A	22
B	17
C	13
D	7
F	2

[8]

From here it is then possible to determine the percentage, or empirical probability, of students who received each grade. By taking the number of students who received each grade and then dividing it by the total number of students in this set the empirical probabilities are as follows:

Percentages of Grades	
A	0.3606557377
B	0.2786885246
C	0.2131147541
D	0.1147540984
F	0.0327868852

[9]

For Group U the sample size is sixteen students, or $n_u = 152$. The grade distribution is:

Grade	Quantity
A	34
B	50
C	33
D	19
F	16

[10]

By the same method as before, the empirical probabilities of receiving each grade are:

Percentages of Grades	
A	0.2236842105
B	0.3289473684
C	0.2171052632
D	0.125
F	0.1052631579

[11]

It is with these values that it is possible to compare the two groups of students.

The method of testing statistical hypotheses is an important element in understanding this study. To begin with there are several initial concepts that are important to understand:

[1] A null hypothesis, denoted by the symbol H_0, which is the hypothesis that we postulate is true.

[2] An alternative hypothesis, denoted by the symbol H_a, which is counter to the null hypothesis and is what we want to support.

[3] A test statistic, calculated from the sample data, which functions as a decision maker.

[4] A rejection region, values of a test statistic for which we reject the null hypothesis and accept the alternative hypothesis.

[A3]

For Group AP, let the probability of earning a particular grade be denoted by $\hat{p}_{ap}$. For Group U, let the probability of earning a particular grade be denoted by $\hat{p}_u$.

In order to determine which data set will best prepare its students for Calculus II, it is best to look at each grade individually and establish a hypothesis test for each grade. The hypothesis tests that will be used in this study are two-tailed tests. In a one-tailed test the alternative hypothesis just considers if one value is greater than or less than the other. In a two-tailed test the alternative hypothesis considers if the two proportions being considered are in fact not equal.

This makes a two-tailed test relevant for this study, in that the only concern is whether or not the Advanced Placement Calculus Exam is equal or not equal to Calculus I.

In these next hypotheses tests normal distribution must be assumed. This is done by the Central Limit Theorem, which states:

If $\overline{X}$ is the mean of a random sample $\overline{X}_1, \overline{X}_2, \ldots, \overline{X}_n$ of size n from a distribution with a finite mean μ and a finite positive variance σ^2, then the distribution of

$$W = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}} = \frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\sigma}$$

is N(0,1) in the limit as $n \to \infty$.

[12]
[A2]

Despite how the populations may look in the histograms, when the sample size, *n*, exceeds that of 25 or 30 the Central Limit Theorem is used in approximation for the distributions. In this particular situation, the total sample size is that of 213. Therefore, for the following cases using normal distribution is appropriate. Also, just as before where $\alpha = 0.10$ thus $\frac{\alpha}{2} = 0.05$ is considered significant for the critical region will be used.

Case A: This case examines the students who receive an A in Calculus II.

$$H_0: p_{ap} = p_u$$

$$H_a: p_{ap} \neq p_u$$

$$z = \frac{|\hat{p}_{ap} - \hat{p}_u|}{\sqrt{\hat{p}(1-\hat{p})(1/n_{ap} + 1/n_u)}} \geq z_{\frac{\alpha}{2}}$$

[13]
[A2]

$\alpha = 0.10$

$\frac{\alpha}{2} = 0.05$

$n_{ap} = 61$

$n_u = 152$

$\hat{p}_{ap} = 0.360656$

$\hat{p}_u = 0.223684211$

$\hat{p} = \frac{22+34}{152+61} = \frac{56}{213} = 0.262911$

Thus, $z = 2.052880618 \geq z_{\frac{\alpha}{2}}$. When $\frac{\alpha}{2} = 0.05,\ z = 1.645$, then $z = 2.052880618 \geq 1.645$ then it is appropriate to reject the null hypothesis.

Case B: This case examines those students who have received a B in Calculus II.

H$_0$: $p_{ap} = p_u$

H$_a$: $p_{ap} \neq p_u$

$$z = \frac{|\hat{p}_{ap} - \hat{p}_u|}{\sqrt{\hat{p}(1-\hat{p})(1/n_{ap} + 1/n_u)}} \geq z_{\frac{\alpha}{2}}$$

[13]
[A2]

$a = 0.10$

$\frac{\alpha}{2} = 0.05$

$n_{ap} = 61$

$n_u = 152$

$\hat{p}_{ap} = 0.278689$

$\hat{p}_u = 0.328947368$

$\hat{p} = \frac{17+50}{152+61} = 0.314554$

Thus, $z = 0.7141197799 \geq z_{\frac{\alpha}{2}}$. When $\frac{\alpha}{2} = 0.05,\ z = 1.645$, then $z = 0.7141197799 \leq 1.645$ then it is appropriate to fail to reject the null hypothesis.

Case C: This case examines those students who have received a C in Calculus II.

H$_0$: $p_{ap} = p_u$

H$_a$: $p_{ap} \neq p_u$

$$z = \frac{|\hat{p}_{ap} - \hat{p}_u|}{\sqrt{\hat{p}(1-\hat{p})(1/n_{ap} + 1/n_u)}} \geq z_{\frac{\alpha}{2}}$$

[13]
[A2]

$a = 0.10$

$\frac{\alpha}{2} = 0.05$

$n_{ap} = 61$

$n_u = 152$

$\hat{p}_{ap} = 0.213115$

$\hat{p}_u = 0.217105263$

$\hat{p} = \frac{13 + 33}{152 + 61} = 0.215962$

Thus, $z = 0.0639795534 \geq z_{\frac{\alpha}{2}}$. When $\frac{\alpha}{2} = 0.05,\ z = 1.645$, then $z = 0.0639795534 \leq 1.645$ then it is appropriate to fail to reject the null hypothesis.

Case D: This case examines those students who have received a D in Calculus II.

$$H_0: p_{ap} = p_u$$

$$H_a: p_{ap} \neq p_u$$

$$z = \frac{|\hat{p}_{ap} - \hat{p}_u|}{\sqrt{\hat{p}(1-\hat{p})(1/n_{ap} + 1/n_u)}} \geq z_{\frac{\alpha}{2}}$$

[13]
[A2]

$a = 0.10$

$\frac{\alpha}{2} = 0.05$

$n_{ap} = 61$

$n_u = 152$

$\hat{p}_{ap} = 0.114754$

$\hat{p}_u = 0.125$

$\hat{p} = \frac{7+19}{152+61} = 0.122066$

Thus, $z = 0.2065013858 \geq z_{\frac{\alpha}{2}}$. When $\frac{\alpha}{2} = 0.05,\ z = 1.645$, then $z = 0.2065013858 \leq 1.645$ then it is appropriate to fail to reject the null hypothesis.

Case F: This case examines those students who have received an F in Calculus II.

H$_0$: $p_{ap} = p_u$

H$_a$: $p_{ap} \neq p_u$

$$z = \frac{|\hat{p}_{ap} - \hat{p}_u|}{\sqrt{\hat{p}(1-\hat{p})(1/n_{ap} + 1/n_u)}} \geq z_{\frac{\alpha}{2}}$$

[13]
[A2]

$\alpha = 0.10$

$\frac{\alpha}{2} = 0.05$

$n_{ap} = 61$

$n_u = 152$

$\hat{p}_{ap} = 0.360656$

$\hat{p}_u = 0.223684211$

$\hat{p} = \frac{2+16}{152+61} = 0.084507$

Thus, $z = 3.249033443 \geq z_{\frac{\alpha}{2}}$. When $\frac{\alpha}{2} = 0.05$, $z = 1.645$, then $z = 3.249033443 \leq 1.645$ then it is appropriate to reject the null hypothesis.

Case Passing: This case examines those students who have received an A, B, or C Calculus II.

H_0: $p_{ap} = p_u$

H_a: $p_{ap} \neq p_u$

$$z = \frac{|\hat{p}_{ap} - \hat{p}_u|}{\sqrt{\hat{p}(1-\hat{p})(1/n_{ap} + 1/n_u)}} \geq z_{\frac{\alpha}{2}}$$

[13]
[A2]

$\alpha = 0.10$

$\frac{\alpha}{2} = 0.05$

$n_{ap} = 61$

$n_u = 152$

$\hat{p}_{ap} = 0.852459$

$\hat{p}_u = 0.7697368$

$\hat{p} = \frac{52+117}{152+61} = 0.79343$

Thus, $z = 1.348127762 \geq z_{\frac{\alpha}{2}}$. When $\frac{\alpha}{2} = 0.05$, $z = 1.645$, then $z = 1.348127762 \leq 1.645$ then it is appropriate to fail to reject the null hypothesis.

Case Failing: This case examines those students who have received a D and F in Calculus II.

$$H_0: p_{ap} = p_u$$

$$H_a: p_{ap} \neq p_u$$

$$z = \frac{|\hat{p}_{ap} - \hat{p}_u|}{\sqrt{\hat{p}(1-\hat{p})(1/n_{ap} + 1/n_u)}} \geq z_{\frac{\alpha}{2}}$$

[13]
[A2]

$\alpha = 0.10$

$\frac{\alpha}{2} = 0.05$

$n_{ap} = 61$

$n_u = 152$

$\hat{p}_{ap} = 0.14754098$

$\hat{p}_u = 0.2302631$

$\hat{p} = \frac{9+35}{152+61} = 0.20657277$

Thus, $z = 1.348119773 \geq z_{\frac{\alpha}{2}}$. When $\frac{\alpha}{2} = 0.05$, $z = 1.645$, then $z = 1.348119773 \leq 1.645$ then it is appropriate to fail to reject the null hypothesis.

In summary, this portion of the study addresses the Advanced Placement Exam versus Calculus I. It can be seen that the samples of the AP Calculus students and Calculus I students come from two different distributions. The nonparametric *U*-test addresses this issue. Based off of the *U*-test the result was to reject the null hypothesis, which states 1.915730804 does not equal 1.645.

This section of the study also addresses hypothesis testing which analyzes the grades of the students. The hypothesis tests relay different results. For a grade of A and F in Calculus II the Advanced Placement Exam rejecting the null hypothesis, $p_{ap} = p_u$, is appropriate. A possible conclusion can include that the Advanced Placement Exam produces a different amount of A's than that of Calculus I. In the cases that discuss grades B, C, and D failing to reject the null hypothesis, $p_{ap} = p_u$, is appropriate. This conclusion leads way to that the grades made in Calculus II by both group of students are possible equal amounts.

Finally, the conclusions of analyzing passing and failing scores of Calculus II are as follows: passing scores include grades of A, B, and C and failing includes grades of D and F. For these hypothesis tests failing to reject the null hypothesis, $p_{ap} = p_u$, is appropriate. Thereby the conclusion is that the Advanced Placement Calculus exam is comparable to that of Calculus I by means of passing and failing.

CHAPTER 5
INTRODUCTION: PRECALCULUS STUDY

At Stetson University there are many different mathematics courses that are offered. Precalculus is one of the more popular mathematics courses taken at Stetson University. As a result, this course is taken by a large range of students; consequently, the caliber of the student's mathematical skills varies. Due to the fact that so large a number of students take Precalculus, apprehension has resulted for a student's readiness in this course.

Precalculus is a beginning level course designed for students who desire to study topics that are commonly found in Calculus. Precalculus topics include exponents, factoring, solving equations and inequalities, graphs, functions, linear and quadratic functions, exponential and logarithmic functions, and systems of linear equations [B3].

In order to be comfortable learning these new concepts students must be able to perform algebraic computations. Students having these skills will help them to be better prepared for Precalculus. It has become apparent that often students are found to struggle with these topics. They do not possess the mathematical skills or background knowledge to be academically successful in this course. Stetson University offers no prerequisite for Precalculus. This allows for many different students to enter this course, resulting in determining if a student is prepared to study the topics that are discussed.

This portion of the study, analyzes which factors of a student may prove influential in a student's success in Precalculus. This research study will analyze different demographics of the students and aide in determining which ones may be most interesting. By the use of multiple linear regression deciding which contributing factors may be most influential for a Precalculus grade.

CHAPTER 6
DATA COLLECTION

The data that was collected for this study was provided by the Stetson University Institutional Research Office. There were many email exchanges with John Tichenor and Patti Sanders. They were able to include efficiently aide in the collection of the data. They were also able to include many of the important demographics of these students.

The data set includes Stetson University students who have completed Precalculus. The contributing factors, or variables, that are used include ethnicity, gender, age, major, class, Stetson GPA, high school GPA, state of residency, SAT Math and Verbal scores, and the grade that was granted for Precalculus. The provided data ranges from 2001-2005 at Stetson University. The sample size of these students consists of 661 records.

Through the course of organizing the data there were many records that were flagged and eliminated. These records did not contribution information. As a result, an approximation of the sample size includes a little over 600 student records. It is then with this sample size that it became important to analyze it.

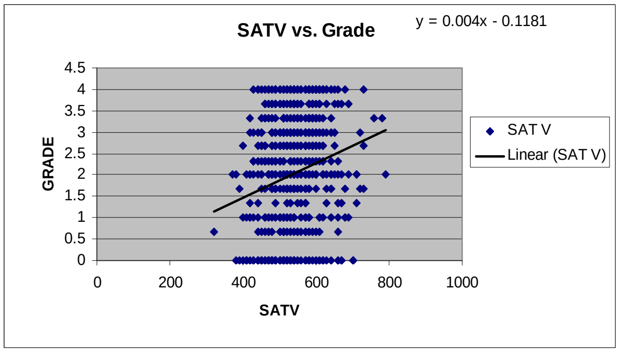
CHAPTER 7
DATA ANALYSIS: REGRESSION

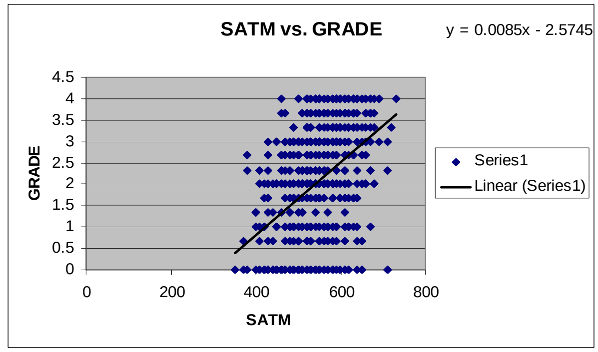
In this next part of the study, the main focus will be on a regression analysis. It will discuss how it can aide in determining which variables are influential for the Precalculus grade. In its simplest form, regression is a method of relating independent and dependent variables to each other. In this study, the dependent variable is that of the Precalculus grade and the independent variables that will be analyzed include those of the different categories provided by the data.

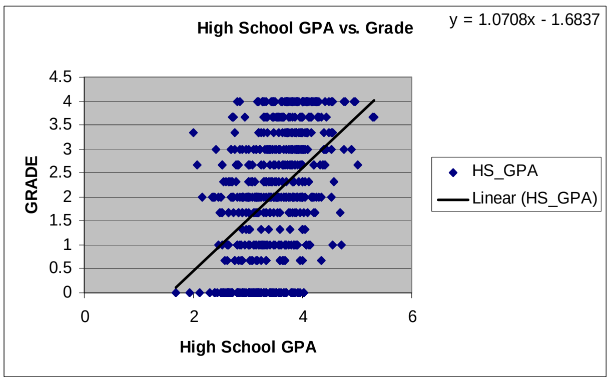
In analyzing the regression, it is important to determine which variables will be used. Of all of the categories that are considered, the ones that will be analyzed include that of SAT Verbal score, SAT Math score, high school GPA, class rank, and age. Through each case the dependent variable will be Precalculus grade. The independent variables will include those previously stated. It is by looking at each of these variables separately and then collectively that conclusions can be made.

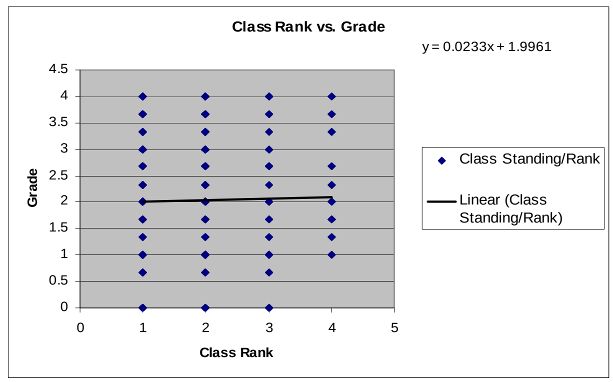
To begin with, the first step in the process would be to look at the relationship between each variable separately versus the Precalculus grade. The following are linear regression models that compare each variable to the Precalculus Grade. These models include the whole data set. These models were created in a spreadsheet where the independent variables are those indicated and the dependent variable is the Precalculus grade.

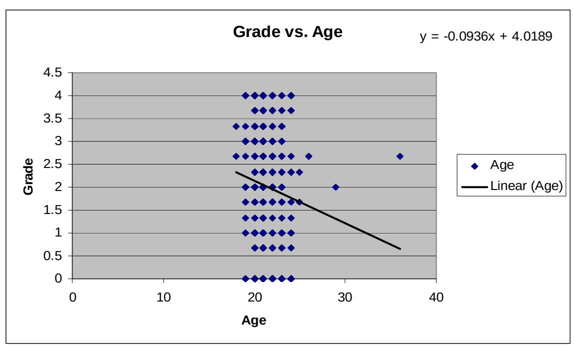
Regression Models:











[14]

After determining the regression for each of the individual variables, the next step is to analyze more than one variable in the regression analysis. Multiple linear regression is the method in which more than one variable is included in a linear regression analysis. Similar to the method of least squares, multiple linear regression is best represented by the equation:

$$\mu(x_1, x_2, ..., x_k) = \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k \text{ with } \beta_1, \beta_2, ..., \beta_k \text{ as linear coefficients}$$
$$\text{with } E(Y) = \mu(x_1, x_2, ..., x_k) \text{ for some observed data,}$$
$$\text{where } E(Y) \text{ is the expected value of the equation } \mu(x_1, x_2, ..., x_k) \text{ [A2].}$$

For this portion of the study, building a linear model is important. It aides in determining what variables are more important than others. A multiple linear regression model is formulated by the above equations. The next step is to determine the linear coefficients, or *β's*. This is done by applying the method of least squares by minimizing:

$$G = \sum_{j=1}^{n} (y_j - \beta_1 x_{1j} - \beta_{2j} - ... - \beta_k x_{kj})^2.$$

Then equate to zero the k first partial derivatives

$$\frac{\partial G}{\partial \beta_i} = \sum_{j=1}^{n} (-2)(y_j - \beta_1 x_{1j} - \beta_2 x_{2j} - ... - \beta_k x_{kj})(x_{ij}),\ i = 1,\ 2,\ ...,\ k$$

Thus, obtained are the k normal equations:

$$\beta_0 n + \beta_1 \sum x_1 + \beta_2 \sum x_2 + ... + \beta_k \sum x_k = \sum y$$
$$\beta_1 \sum x_1^2 + \beta_2 \sum x_1 x_2 + ... + \beta_k \sum x_1 x_k = \sum x_1 y,$$
$$\beta_1 \sum x_2 x_1 + \beta_2 \sum x_2^2 + ... + \beta_k \sum x_2 x_k = \sum x_2 y,$$
$$\beta_1 \sum x_k x_1 + \beta_2 \sum x_k x_2 + ... + \beta_k \sum x_k^2 = \sum x_k y$$

$$\text{with } \sum_{i=1}^{n} x_{i1} \text{ as } \sum x_1, \sum x_{1i} x_{2i} \text{ as } \sum x_1 x_2 \text{, and so on. [A2][A3]}$$

These normal equations provide the values for the linear coefficients. It is with these equations that it is possible to compute the regression models for the set of variables. The following is an example of multiple linear regression. All of the multiple linear regression models were constructed by this method. The following is an example of one of the multiple linear regression models. This example shows SAT Math Score and SAT Verbal Score versus Precalculus Grade.

Example: SAT Math Score, SAT Verbal Score vs. Grade

Computing the values in a spreadsheet the following values will be used:

with $x_1 = \text{SAT Verbal Score}$, $x_2 = \text{SAT Math Score}$, and $y = \text{Precalculus Grade}$.

$n = 647$

$\sum x_1 = 348680$

$\sum x_2 = 351160$

$\sum y = 1315.54$

$\sum x_1^2 = 191031800$

$\sum x_1 x_2 = 190630500$

$\sum x_2^2 = 193664400$

$\sum x_1 y = 721430.8$

$\sum x_2 y = 740090.5$

The normal equations are as follows:

$$\beta_0 n + \beta_1 \sum x_1 + \beta_2 \sum x_2 = \sum y$$
$$\beta_0 \sum x_1 + \beta_1 \sum x_1^2 + \beta_2 \sum x_1 x_2 = \sum x_1 y$$
$$\beta_0 \sum x_2 + \beta_1 \sum x_1 x_2 + \beta_2 \sum x_2^2 = \sum x_2 y$$

Then by substituting in the values obtained are three equations with three unknowns:

$$647\beta_0 + 348680\beta_1 + 351160\beta_2 = 1315.54$$
$$348680\beta_0 + 191031800\beta_1 + 190630500\beta_2 = 721430.8$$
$$351160\beta_0 + 190630500\beta_1 + 193664400\beta_2 = 740090.5$$

By solving these equations the coefficients can be determined. This was done by substituting the values into a matrix and then computing this via reduced-row echelon form:

$$\begin{bmatrix} 647 & 348680 & 351160 & 1315.54 \\ 348680 & 191031800 & 190630500 & 721430.8 \\ 351160 & 190630500 & 193664400 & 740090.5 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & -2.658477789 \\ 0 & 1 & 0 & 2.851935194 \times 10^{-4} \\ 0 & 0 & 1 & .0083612423 \end{bmatrix}$$

Thereby making the following values for the coefficients:

$\beta_0 = -2.658477789$
$\beta_1 = 2.851935194 \times 10^{-4}$
$\beta_2 = .0083612423$

Once the coefficients have been determined, by the method of least squares, the following equation is evident:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

Substituting in the values of the coefficients the equation that fits this data:

$$y = -2.658477789 + 2.851935194 \times 10^{-4} x_1 + .0083612423 x_2 .$$

This equation represents the multiple linear regression model of a Precalculus grade to the observed variables of the SAT Verbal Score and the SAT Math Score. [15]

By applying this same method, the following are the equations for the different sets of variables:

SAT Math Score, Class, vs. Grade:

$y = -2.676601871 + .0085158631 x_1 + .0552144765 x_2$

With $x_1 =$ SAT Math Score, $x_2 =$ Class, and $y =$ Precalculus Grade

SAT Math Score, Age, vs. Grade:

$y = -1.168970644 + .0083926301 x_1 + -.0637596142 x_2$

With $x_1 =$ SAT Math Score, $x_2 =$ Age, and $y =$ Precalculus Grade

SAT Math Score, High School GPA, vs. Grade:

$y = -4.722402413 + .0068274188x_1 + .8782964774x_2$

With x_1 = SAT Math Score, x_2 = High School GPA, and y = Precalculus Grade

SAT Math Score, SAT Verbal Score, High School GPA vs. Grade:

$y = .1153632706 + .0148997052x_1 + 1.135427505x_2 + -.0187942803x_3$

with

x_1 = SAT Math Score, x_2 = High School GPA, x_3 = SAT Verbal Score, and y = Precalculus Grade

SAT Math Score, High School GPA, Age vs. Grade:

$y = -3.784374054 + .0070501747x_1 + .7903791953x_2 + -.0352491028x_3$

With x_1 = SAT Math Score, x_2 = High School GPA, x_3 = Age, and y = Precalculus Grade

All Categories:

$$y = -3.738078388 + -0.0360148259x_1 + 0.0562235746x_2 + \\ -6.959682722\times10^{-4}x_3 + 0.0072671597x_4 + 0.8292558429x_5$$

With x_1 = Age, x_2 = Class, x_3 = SAT Verbal Score, x_4 = SAT Math Score, x_5 = High School GPA, and y = Precalculus Grade.

Once these equations were determined the next step in the process was to look at different forms of multiple linear regression equations. The following equation addresses when a student's High School GPA and SAT Math Score are both squared. These computed values can still be viewed as linear variables. To illustrate, if x_1=SAT Math Score, then x_1^2= SAT Math Score squared; then by letting $x_1^2 = x_2$ a linear model can be constructed.

It is by this procedure that it is possible to analyze the next set of regression models.

High School GPA^2, SAT Math Score^2 vs. Precalculus Grade:

$y = 2.208780908 + -6.947830637\times10^{-8}x_1 + 0.0033391723x_2$

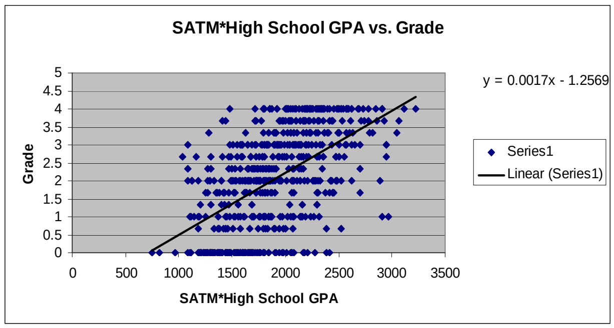
With x_1 = SAT Math Score^2, x_2 = High School GPA^2, and y = Precalculus Grade

High School GPA^2, SAT Math Score vs. Precalculus Grade:

$$y = -3.206651599 + 0.0068122235x_1 + 0.1246346659x_2$$

With x_1 = SAT Math Score, x_2 = High School GPA^2, and y = Precalculus Grade

This next regression model was computed by making the variable to be considered the product of the SAT Math Score and the High School GPA.



[14]

Finally, the last set of multiple regression equations analyzes the possibility of creating quadratic regression equations. As it will be later shown, the two variables that appear to be the most influential are that of SAT Math Score and High School GPA. It is by this that examining these two variables by quadratic multiple regression could prove to be interesting. Similarly as before, the following are the normal equations that satisfy for quadratic equations with one variable:

$$\begin{aligned} \beta_0 n + \beta_1 \sum x_i + \beta_2 \sum x_i^2 &= \sum y_i \\ \beta_0 \sum x_i + \beta_1 \sum x_i^2 + \beta_2 \sum x_i^3 &= \sum x_i y_i \quad \text{[A2].} \\ \beta_0 \sum x_i^2 + \beta_1 \sum x_i^3 + \beta_3 \sum x_i^4 &= \sum x_i^2 y_i \end{aligned}$$

The following quadratic equations are:

High School GPA vs. Precalculus Grade:

$$y = 2.174651972 + -1.15301758x + 0.3125883394x^2$$

With x = High School GPA and y = Precalculus Grade

SAT Math Score vs. Precalculus Grade:

$$y = -6.889744762 + 0.0246024386x + -1.47997643 \times 10^{-5} x^2$$

With x = SAT Math Score and y = Precalculus Grade.

In the following section, these regression models will prove to be more useful.

CHAPTER 8
DATA ANALYSIS: RESIDUALS

Not much can be determined by just constructing a regression model. There needs to be a method to establish the accuracy of these models. Residuals assist in demonstrating if a given model is a good fit to the observed data. A form of residuals is the R^2 statistic. This study will use the R^2 statistic to show which variables may be more influential for a Precalculus Grade. "The R^2 statistic measures the amount of variation in the data that is explained by a model," [A5]. The R^2 statistic follows the formula:

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}, \text{ where } SSR = \sum(\hat{y}_i - \bar{y})^2,\ SSE = \sum(y_i - \hat{y}_i)^2, \text{ and } SST = SSR + SSE.$$

Where $\hat{y}$ is the model or the predicted value of y, $\hat{y}$ is the mean of y, and y is the observed data [A3].

In order to determine these sums, it was efficient to enter the information into a spreadsheet and have it compute the sums of: each independent variable, the dependent variable of y, y-predicted value, error between y and ypredicted, error^2, y-ymean, (y-ymean)^2, ypredicted-ymean, and (ypredicted-ymean)^2.

Microsoft Excel determined the first set of R2 values of each individual variable. They are as follows:

SAT Verbal Score vs. Precalculus Grade: $R^2 = 0.0426$

SAT Math Score vs. Precalculus Grade: $R^2 = 0.1895$

High School GPA vs. Precalculus Grade: $R^2 = 0.2003$

Class Rank vs. Precalculus Grade: $R^2 = 0.0002$

Age vs. Precalculus Grade: $R^2 = 0.0089$

For the rest of the multiple regression models, determining the R^2 statistic was determined by the formula listed above. An example of this is as follows:

Example of R^2: SAT Verbal Score, SAT Math Score vs. Precalculus Grade

	x1	x2	y	y predicted	error	error^2	y-ymean	(y-ymean)^2	ypred-ymean	(ypred-ymean)^2
	SAT_V	SAT_M	GRADE							
	580	650	4	2.941741947	1.058258053	1.119910106	1.966707883	3.867939895	0.90844983	0.825281093
Mean of y:	530	620	4	2.676645002	1.323354998	1.75126845	1.966707883	3.867939895	0.643352885	0.413902934
2.033292	620	580	4	2.367862727	1.632137273	2.663872078	1.966707883	3.867939895	0.33457061	0.111937493
	570	600	4	2.520827897	1.479172103	2.18795011	1.966707883	3.867939895	0.48753578	0.237691136
	540	540	4	2.010597553	1.989402447	3.957722094	1.966707883	3.867939895	-0.022694564	0.000515043
	590	540	4	2.024857229	1.975142771	3.901188964	1.966707883	3.867939895	-0.008434888	7.11473E-05
	540	540	4	2.010597553	1.989402447	3.957722094	1.966707883	3.867939895	-0.022694564	0.000515043
	600	520	4	1.860484319	2.139515681	4.577527351	1.966707883	3.867939895	-0.172807799	0.029862535
	490	520	4	1.829113032	2.170886968	4.71275023	1.966707883	3.867939895	-0.204179086	0.041689099
	570	640	4	2.855277589	1.144722411	1.310389398	1.966707883	3.867939895	0.821985472	0.675660116
	550	570	4	2.264286758	1.735713242	3.01270046	1.966707883	3.867939895	0.23099464	0.053358524
	620	620	4	2.702312419	1.297687581	1.683993058	1.966707883	3.867939895	0.669020302	0.447588164
	630	690	4	3.290451315	0.709548685	0.503459336	1.966707883	3.867939895	1.257159198	1.580449249
	580	590	4	2.440067409	1.559932591	2.433389688	1.966707883	3.867939895	0.406775292	0.165466138
	520	460	4	1.335994299	2.664005701	7.096926374	1.966707883	3.867939895	-0.697297818	0.486224248
	600	730	4	3.616345202	0.383654798	0.147191004	1.966707883	3.867939895	1.583053084	2.506057067
	580	500	4	1.687555602	2.312444398	5.347399093	1.966707883	3.867939895	-0.345736515	0.119533738
	620	690	4	3.28759938	0.71240062	0.507514643	1.966707883	3.867939895	1.254307263	1.573286709
	600	550	4	2.111321588	1.888678412	3.567106145	1.966707883	3.867939895	0.07802947	0.006088598
	490	540	4	1.996337878	2.003662122	4.014661901	1.966707883	3.867939895	-0.03695424	0.001365616
	600	600	4	2.529383703	1.470616297	2.162712294	1.966707883	3.867939895	0.496091585	0.246106861
	580	610	4	2.607292255	1.392707745	1.939634862	1.966707883	3.867939895	0.574000138	0.329476158
	730	680	4	3.235358244	0.764641756	0.584677015	1.966707883	3.867939895	1.202066127	1.444962973
	580	550	4	2.105617717	1.894382283	3.588684233	1.966707883	3.867939895	0.0723256	0.005230992
	560	670	4	3.103262923	0.896737077	0.804137386	1.966707883	3.867939895	1.069970805	1.144837524
	610	520	4	1.863336254	2.136663746	4.565331964	1.966707883	3.867939895	-0.169955864	0.028884996

*This is a sample from the spreadsheet that was used to calculate the sums. [16]

Then the resulting sums are as follows and also applying the formula for R^2:

SSE	946.9864723
SSR	221.6124144
SST	1168.598887
R^2	0.189639419

By using the same method as the example the following R^2 values are: [17]

SAT Math Score, Class vs. Precalculus Grade

SSE	946.0080481
SSR	222.5908377
SST	1168.598886
R^2	0.190476681

SAT Math Score, Age vs. Precalculus Grade

SSE	942.4042
SSR	225.6911
SST	1168.095
R^2	0.193213

SAT Math Score, High School GPA vs. Precalculus Grade

SSE	756.4775341
SSR	349.0783151
SST	1105.555849
R^2	0.315749146

SAT Math Score, SAT Verbal Score, High School GPA vs. Precalculus Grade

SSE	1494.441536
SSR	1141.580849
SST	2636.022385
R^2	0.433069482

High School GPA Quadratic Regression vs. Precalculus Grade

SSE	893.2379914
SSR	213.5427614
SST	1106.780753
R^2	0.192940436

SAT Math Score Quadratic Regression vs. Precalculus Grade

SSE	941.9254985
SSR	226.5047293
SST	1168.430228
R^2	0.193853877

SAT Math Score^2, High School GPA^2 vs. Precalculus Grade

SSE	1117.857
SSR	21.43976
SST	1139.297
R^2	0.018818

Combined Categories vs. Precalculus Grade

SSE	758.3127175
SSR	347.3834531
SST	1105.696171
R^2	0.314176229

In order to interpret when R^2 statistic is the best it must be the value that is the closet to one [A5]. Applying this, the regression models from the largest to the smallest R^2 values are:

1. SAT Math Score, SAT Verbal Score, High School GPA vs. Precalculus Grade
2. SAT Math Score, High School GPA vs. Precalculus Grade
3. Combined Categories vs. Precalculus Grade
4. High School GPA vs. Precalculus Grade
5. SAT Math Score Quadratic Regression vs. Precalculus Grade
6. SAT Math Score, Age vs. Precalculus Grade
7. High School GPA Quadratic Regression vs. Precalculus Grade
8. SAT Math Score, Class Rank vs. Precalculus Grade
9. SAT Verbal Score, SAT Math Score vs. Precalculus Grade
10. SAT Math Score vs. Precalculus Grade
11. SAT Verbal Score vs. Precalculus Grade
12. SAT Math Score^2, High School GPA^2 vs. Precalculus Grade
13. Age vs. Precalculus Grade
14. Class Rank vs. Precalculus Grade

The regression model that had the highest R^2 value was that of SAT Math Score, SAT Verbal Score, and High School GPA. The two individual variables that produce the highest R^2 values include that of SAT Math Score and High School GPA. There are some possible reasons as to why these two individual variables are higher than the others. A SAT Math Score could measure a student's ability in mathematics and their High School GPA could represent their work ethic, or how hard a student works in his or her classes.

CHAPTER 9
CONCLUSION

This study addresses two different topics. The first study addresses the Advanced Placement Calculus Exam and how it compares to that of Calculus I. The second study discusses regression of Precalculus and how it can help to determine which contributing factors may be interesting to a Precalculus grade.

In the Advanced Placement Calculus Exam portion of this study two groups are determined, those of AP students and those of Calculus I students at Stetson University and the University of Central in Florida. In this study two different results were found. One is both groups come from two different distributions. This was completed by a nonparametric U-test. The *U*-test was able to take the given information and determine that the two groups were not of the same distributions. In addition to the distributions of the groups, hypothesis tests were also constructed that analyzed the grades that were made by those students in Calculus II from the two different groups. The hypothesis tests that were created were two-tailed tests. They had a null hypothesis that stated the empirical probabilities of these two groups were in fact equal and an alternative hypothesis that stated that the empirical probabilities of these two groups were not equal. The conclusions for these hypothesis tests were that for grades of an A and F in Calculus II it was appropriate to reject the null hypothesis and for grades B, C, and D in Calculus II failing to reject the null hypothesis was valid.

The second portion of this study examines a regression of Precalculus students at Stetson University. The main focus of this section is to determine which factors may be more influential in a student's Precalculus grade. The different variables that were considered for the regression models included that of age, class rank, High School GPA, SAT Math and Verbal Scores, and the grade that was given for Precalculus. Regression models were created for each of these variables separately. After the variables were considered individually, studying the different combinations of these variables to build regression models was the next step.

Once the regression models were built, it became necessary to determine which variables would be more prominent for a Precalculus grade. This was done by computing an R^2 statistic. The R^2 statistic justifies what variables may be a good fit to the model. The two individual variables that were the most prominent included that of SAT Math Score and High School GPA. The multiple regression model with the highest R^2 value was that of SAT Math Score, SAT Verbal Score, and High School GPA versus Precalculus Grade.

REFERENCES

A. OFFLINE

[A1] H. Adler and E. Roessler, *Introduction to Probability and Statistics: Fourth Edition,* W.H. Freeman and Company. San Francisco, CA, 1968.

[A2] R. Hogg and E. Tanis, *Probability and Statistical Inference: Sixth Edition*, Prentice-Hall, Inc. Upper Saddle River, NJ, 2001.

[A3] W. Mendenhall and T. Sincich, *A Second Course In Statistics: Regression Analysis: Sixth Edition*, Pearson Education, Inc. Upper Saddle River, NJ, 2003.

[A4] I. Miller and M. Miller, *John E. Freund's Mathematical Statistics: Sixth Edition,* Prentice-Hall, Inc. Upper Saddle River, NJ, 1999.

[A5] D. Mooney and R. Swift, *A Course in Mathematical Modeling,* The Mathematical Association of America Inc. United States, 1999.

B. ONLINE

[B1] http://www.edonline.com/collegecompass/s_test.htm - Educational On-Line: Specializing in Internet Interface Applications that Centralize and Present Information for Educational Institutions Web Site.

[B2] http://apcentral.collegeboard.com/exam/0,,152-0-0-0,00.html – The Official College Board AP Central Webpage for Educators.

[B3] http://www.stetson.edu/bulletin/extras/artsci.pdf - Stetson University Bulletin: College of Arts and Sciences.

[B4] http://www.stat.berkeley.edu/users/stark/SticiGui/Text/gloss.htm - Glossary of Statistical Terms.