Assessment of Present Cal Poly Admission Requirements

Scott Goodell
Advisor: Dr. Steve Rein

Table of Contents

I. Research Goal

Being accepted into college is one of the most fulfilling times in one's life. The idea of starting a new chapter in life, usually in a complete distinct destination as the present, while advancing your education and knowledge is something to be excited about. During a college application process, people often wonder how schools choose which students to accept and which to deny. There are many factors that contribute to a student's acceptance into a university. Every school has a different set of criterion when looking at students and deciding who they will accept for the upcoming year. For my senior project, I have chosen to use my statistical knowledge to compare and assess the current admission criteria for California Polytechnic State University, San Luis Obispo. My overall goal, using the given data and appropriate techniques, is to find ways in which the criteria is sufficiently viewing students applications and ways that it could possibly be looked at differently or improved.

II. Multiple Criteria for Admissions

Cal Poly's current admission process for the 2010-2011 years follow a point procedure called the Multi-Criteria Admissions for Freshman applicants. There are four main sections that are weighted to determine if a student will be admitted to the university. The four categories are:

1. Academic Performance (GPA)
2. Program of Study with respect to College Preparatory Coursework.
3. Standardized Test Scores (SAT)
4. Extracurricular Activities

Section 1: Academic Performance

A student's main high school coursework GPA accounts for a weight between 45% and 55% towards the students overall final admission score. Cal Poly accounts for students taking Honors and AP classes by allowing for a cumulative GPA of over a 4.00. For example, a student can apply to Cal Poly with a high school GPA of 4.25. I believe everyone feels that high school GPA should be one of the strongest indicators of whether or not a student gets accepted into a university.

Section 2: MCA Scoring of CSU Required Courses

During high school, a student is required to take a certain amount of semesters of required courses that results in 10% - 19% of the overall points for an admission score. Cal Poly requires a minimum of 28 semesters of required coursework plus at least 2 semesters of elective classes. Required courses include:

- English (8 Semesters)
- Algebra (4)
- Geometry (2)
- Foreign Language (4)
- Lab Science (4)
- History (4)
- Performing Arts (2)

This totals a minimum of 28 semesters that allows for a student to take 2 semesters of elective classes. An applying student is encouraged to take more than the minimum required semesters and is awarded bonus points for doing so. A student who takes the highest possible number of semesters and gets the maximum amount of bonus points is presented below.

- English (10 Semesters; 100 Bonus Points)
- Algebra (4 ; 0 Bonus Points)
- Geometry (2 ; 0 Bonus Points)
- Foreign Language (8 ; 100 Bonus Points)
- Lab Science (8 ; 200 Bonus Points)
- History (4 ; 0 Bonus Points)
- Performing Arts (4 ; 50 Bonus Points)
- Advanced Math (4 ; 500 Bonus Points)

As seen above, this student has received a maximum of 950 admission bonus points for their high school curriculum. Notice that extra semesters of Lab Sciences and Advanced Mathematics have a higher weight than the rest of the subjects with respect to the allotment of Bonus points. A student taking 4 extra semesters of Advanced Math results in 52% of the maximum possible bonus points by itself. For each semester that a student is under the minimum required number of courses, a 500 point deduction of admission score is applied for each of the semesters under the required 30 semesters.

Section 3: Standardized Test Scores

A student's SAT score accounts for 25% to 35% of the total admission score. Cal Poly looks only at the SAT Math and Verbal section. The SAT Writing section is not considered with students applications. It should be noted that equal weights have been placed on both Math and Verbal sections of the SAT. Equal points are awarded for how well someone performs on each section. The maximum number of points that someone can earn is 825 per section, or 1650 total.

Section 4: Work Experience and Extra-Curricular Activities

A student's work experience and extra-curricular activities account for a weight between 1% and 10% of the total admission score.

Each college establishes a minimum requirement score for acceptance for each student that applies. A college also must roughly decide how many students that will be accepted overall. The criterion for admission recommends the score be based on a student’s probability of obtaining at least a “C” in his or her studies at Cal Poly. If a given application meets the required minimum score, then the student has reason to be admitted into the college where they had applied. If there is an overflow of students that meet the minimum criterion score set by each college, bonus points are awarded for particular qualities of a student. These qualities include:

- California recently released veteran
- Service Area
- Partner Schools
- Father’s highest education level: No/Some High School
- Mother’s highest education level: No/Some High School
- Faculty or Staff Dependent

Any of these qualities for a student listed above can boost the student’s overall admission score, contributing to a higher probability of the student being accepted to Cal Poly. This method of admission criteria is designed to adequately weight every student’s qualities in order to admit students with the highest probability of succeeding in their studies.

III. Data

The data that was used for this project is a list of admitted freshman students from Fall 2006 to Fall 2010. For this study, I will only be using the students that were admitted freshmen in Fall 2006. For each observation, there is a long list of variables that describe the high school and present and past undergraduate statistics for each student. The list of variables include:

- Admitted Term
- Observation ID
- Degree Type
- College Code
- Major Code
- Gender
- Ethnicity Code
- Residency Code
- Area of Admissions Grouping
- Reason Code for Time of Admission
- High School GPA
- High School Rank
- SAT Reading
- SAT Math
- Each ACT Score (5 Sections)
- Geographic Region
- Parent Education
- GPA for each individual term from 2006 to 2009
- Cumulative CPA for each individual term from 2006 to 2009
- Degree Term

Based on the MCA for freshmen students, there are a few ideas that are not listed in the dataset. There are no variables that designate if a student that applied had the necessary CSU class requirements or any of the extra bonus semesters for the classes specified earlier. Another idea that is not in the dataset is the specified number of extracurricular activities and the work experience that each student had when applying.

The goal of the analysis is to use variables from the data set to find ways in which the MCA is sufficiently evaluating a student’s applications as well as finding ways the current admission criterion could be altered or looked at differently. The implications of this study include possible rethinking of the way students are accepted into Cal Poly.

IV. Methods

Based on my research goal, I am hoping to find a way to interpret success for a student and decide if the MCA is correctly admitting students, based on their qualifications and statistics in order to maximize the number of students that will hopefully succeed in their studies. At the moment, the goal of the MCA is to find students that they believe have the highest chance of earning a grade of, at least, a "C" in each of their classes while at Cal Poly. This is a similar criterion that I will use to assess success for a student during their time of study. To accomplish this goal I defined a variable called "Verdict" in the data set that each student will receive. To build this model assessing success, I will be using Binary Logistic Regression as my analysis procedure. The response variable for this model will be the created "Verdict" variable for each of the students. This variable has two levels, S for Success and F for Fail and the analysis chosen will create a model that predicts the probability of a given student succeeding or failing if admitted to Cal Poly given each of the explanatory variables included in the model.

$$\ln\left(\frac{\pi(x)}{1-\pi(x)}\right) = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k$$

For this project, my definition of success is a student that has been at Cal Poly for over two years and has maintained at least a 2.3 GPA by fall of their third year. A 2.3 GPA equates to a student averaging a C+ in all of their classes. The two full years of school implies that a student has not transferred from the school or has not yet failed out of Cal Poly. The year constraint is to make sure that the student has established their GPA over a respectable period of time. So a student that has at least a 2.3 GPA after fall of their third year and is still at Cal Poly is considered a Success. If a student does not meet either of these success requirements then they are considered a failure. This verdict variable can now be predicted using a combination of explanatory variables listed earlier in the section. The model used will predict the probability of a student being a success.

Model:

With the list of explanatory variables above, I have decided the variables that I feel may influence a student's probability of success are as follows.

1. Gender
2. Ethnicity Code
3. High School GPA
4. SAT Reading
5. SAT Math
6. ACT

Manipulation:

To conduct my model, there were alterations to the data that needed to be made.

The verdict variable was discussed as giving a student either success or failure based on having at least a 2.3 GPA and being at Cal Poly for two full years. The GPA of a student was measured using the cumulative GPA of Fall 2009. To count the number of quarters, an indicator of 1 was created for each of the quarters that a student had a recorded GPA in the dataset. I added all of these quarters together to create a total quarter's variable that measured how many quarters a student had studied at Cal Poly. If a student had satisfied both requirements listed above, then the Verdict variable was codes as S. If either of the criteria above was failed, then Verdict was coded as F.

For the variable Ethnicity Code, the data had numerical values for each of the students. I renamed each of the numbers with their actual name given in the explanation of the data set and made the variable categorical classes. There are six total levels for Ethnicity:

- Asian
- African-American
- Hispanic
- Native American
- White
- Other

For the variable ACT, I simply checked to see if a student had a score for the ACT math category. If the student had taken the ACT then the value was Y, for yes, otherwise it was coded as N, for no. The reason for this variable is because I believe that included the scores from both standardized tests, SAT and ACT, would create issues of multicollinearity. I did, however, want to try and discover if the rate of success changed for a student that had taken both the SAT and ACT as oppose to a student that simply took the SAT.

A variable called Category was create in order to put choice students into particular school type groups that could be used as predictors in one of the choice models. The six possible values for Category are Agriculture, Science, Engineering, Liberal Arts, Mathematics, and Business. The allocation of students in each Category is presented in Table 1 below:

Table 1: Allocation of Majors to Categories

Category	**Student's Major(s)**
Agriculture	Recreation, Nutrition, Environmental Studies, Horticulture, Earth Science, and Food Science.
Science	Biology
Engineering	Electrical Engineering
Liberal Arts	Art, Communications, History, English, Journalism, Modern Languages and Literature, Music, and Political Science
Mathematics	Mathematics, Statistics, and Physics
Business	Business Administration

The choices for each major were done to get homogeneity of student characteristics within a category and to get a relatively equal sample size in each of the Category groups.

Based on these changes, the variables of interest will be used to predict the odds of success. There will be eight total logistic regression procedures, each using the binary variable Verdict as the response variable. The eight models can be separated into three sub categories.

1. Overall model for the entire 2006 Fall Freshmen class. This model will include each of the six variables of interest as explanatory variables.

2. Category including six models that target a specific Cal Poly major. Each of these models also includes each of the six explanatory variables of interest. The six specific majors of interest are listed. The specific students that are associated with each model are the students associated with the specific majors in Table 1.

    a. Agriculture
    b. Science
    c. Engineering
    d. Mathematics
    e. Liberal Arts
    f. Business

3. This model is trying to understand if students tend to have different chances of success based on quantitative variables and what area of study the students come from. The model includes the six variables of interest and includes the Category variable and well as three interaction terms: SAT Math by Category, SAT Reading by Category, and High School GPA by Category.

Theses eight total procedures have been my choice of study for this project, we now proceed with the analysis of the data.

V. Descriptive Statistics

Before any tests are run, descriptive statistics of the sample are shown for convenience.

Table 2: Data Demographics by Gender

Gender	Count	Percentage	%Success
Male	2079	56.68%	66.33%
Female	1589	43.32%	74.26%

Table 2 above, shows the distribution of the data with respect to Gender. One can notice that the number of males is larger than the number of females. However, the percentage of the Gender that was deemed successful is larger for the females then the males. Of all the females in the data set, almost 75% were deemed successful during their undergraduate career. For the males, about 66% of the students were deemed successful.

Figure 1: Proportion of Success by Gender

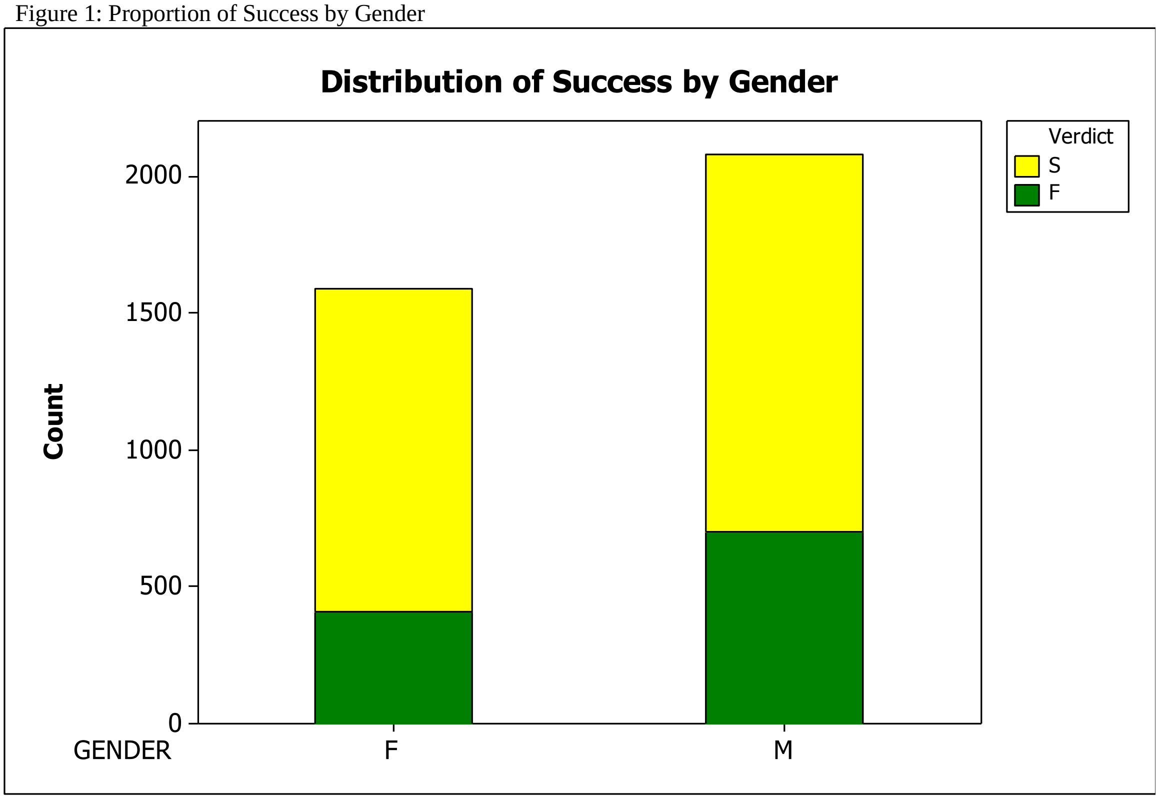


Figure 1 above shows the distribution of Success and Failure with respect to each Gender. As shown above, females seem to have a higher proportion of Successes than the males, even though males clearly outnumber the females for the 2006 Fall Freshmen class.

Table 3: Data Demographics by Ethnicity

Ethnicity	Count	Percentage	%Success
Asian	448	12.21%	63.62%
African-American	51	1.39%	56.86%
Hispanic	379	10.33%	59.37%
Native American	23	0.63%	65.22%
White	2391	65.19%	72.86%
Other	376	10.25%	69.95%

Table 3 shows the distribution of the data with respect to the six Ethnicity levels listed earlier in the methods section of the report. Whites outnumber any of the ethnicities and also have the highest proportion deemed successful. African-Americans have the lowest proportion of successful students out of any of the ethnicities. However, each of the ethnicities have around the same proportion deemed successful. There are small counts of Native Americans and African-Americans in the table.

Figure 2: Distribution of Success by Ethnicity using Cumulative Percentage

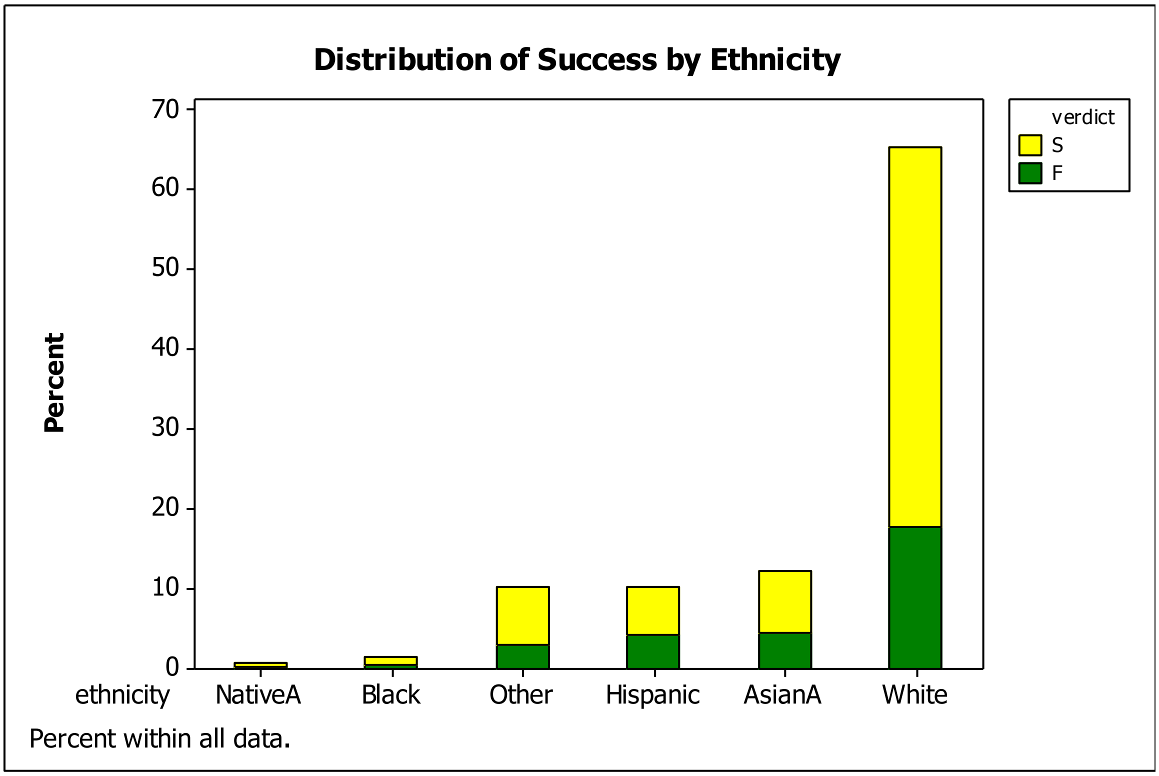


Figure 2 corresponds with the statistics shown in Table 3 above. It is difficult to tell from the graph if any Ethnicity stands alone with the highest proportion of success.

Table 4: Descriptive Statistics for Quantitative Response Variables

Variable	N	Mean	StDev	Min	Max
HSGPA	3388	3.70	0.38	2.00	4.68
SAT Math	3552	615.07	78.84	310	800
SAT Reading	3552	569.86	75.18	310	800

Table 4 shows the descriptive statistics for the quantitative variables that have been included in the model. It is shown that the number of High School GPA values in the data set is less than the number of SAT Math and SAT Reading values, which may seem odd. The reason for this is because there were 164 students that did not have a High School GPA value. From the table, it shows that the 2006 Fall Freshmen class sported an impressive 3.70 average High School GPA. The average SAT Math score was higher than the average SAT Reading score by a healthy amount. Later, a distribution of these three choice variables by the students Category will be produced. That table will give rise to possibilities of differences in skills based on student background.

Figure 3: Histogram of High School GPA

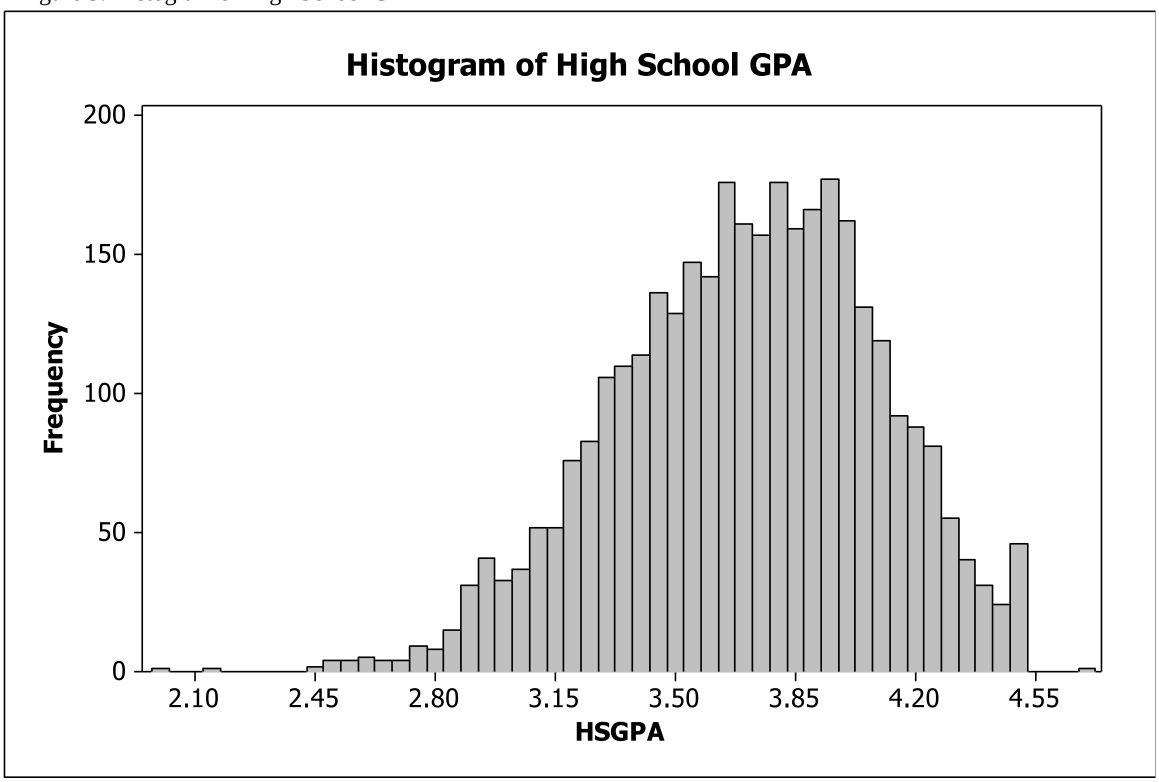


Figure 3 is a histogram of High School GPA for all students in the 2006 Fall Freshmen class. The distribution seems to be fairly normal; but a bit skewed to the left. This is most likely due to the fact that it is rare that a student with such a low GPA will ever be admitted to Cal Poly.

Figure 4: Histogram of SAT Math

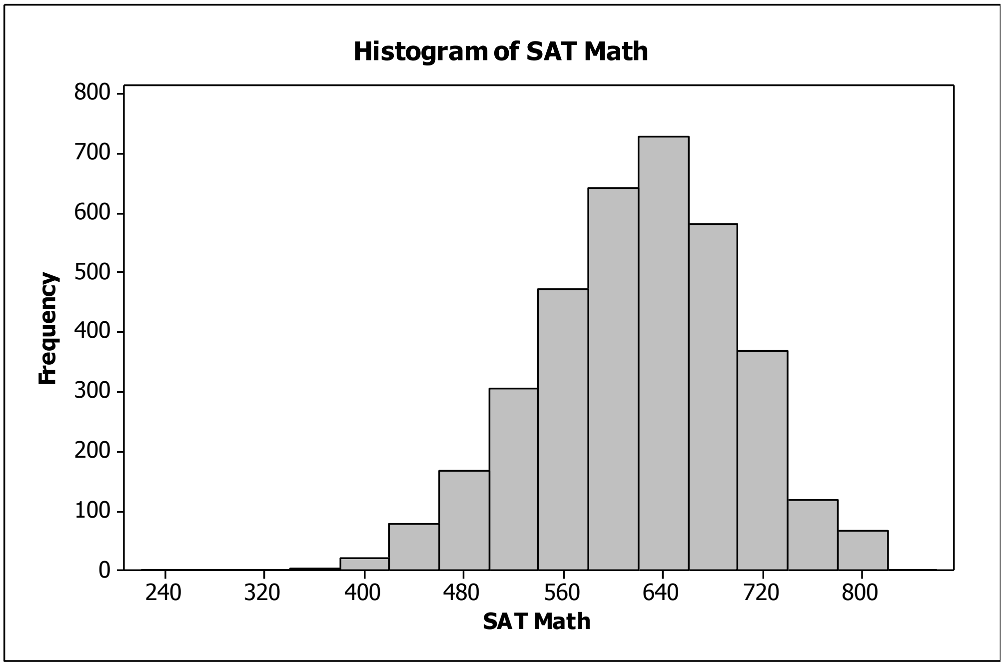
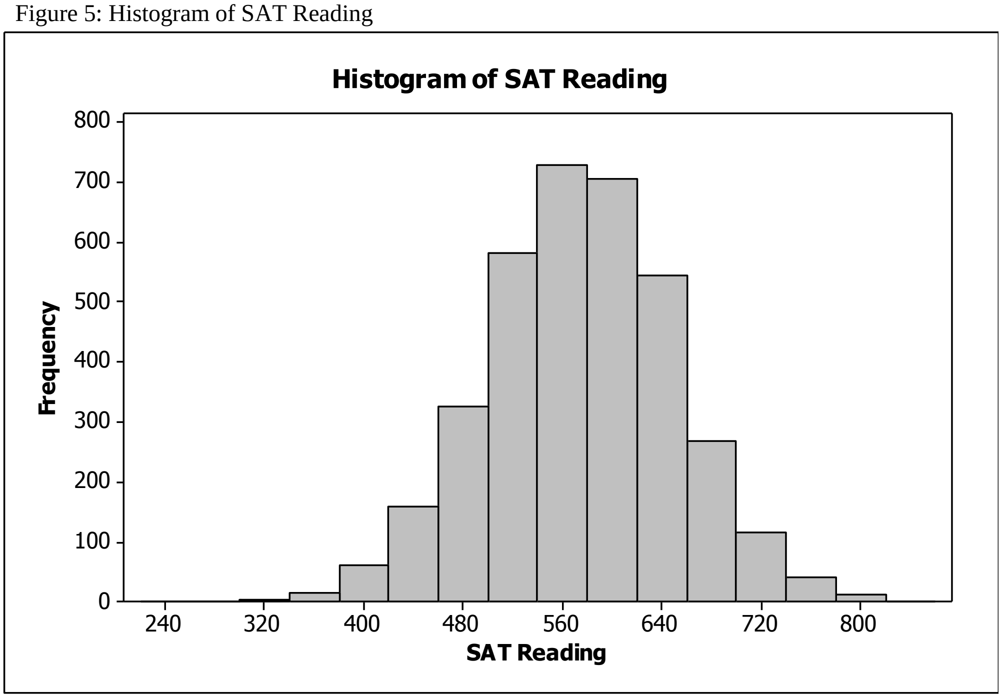


Figure 5: Histogram of SAT Reading



Figures 4 and 5 are histograms for SAT Math and SAT Reading scores respectively for the freshmen class. Both of the distributions look normal with the SAT Math score being slightly skewed left. The spikes in the bars for each of the graphs are interesting to look at. It is possible that there are scores on the tests that are obtained with more combinations of right and wrong answers. It seems on average that a student in high school applies with a higher SAT Math score than an SAT Reading score. It should be noted that these are the values only for students that were admitted, not for all students that applied.

Figure 6: Scatterplot between SAT Math and SAT Reading

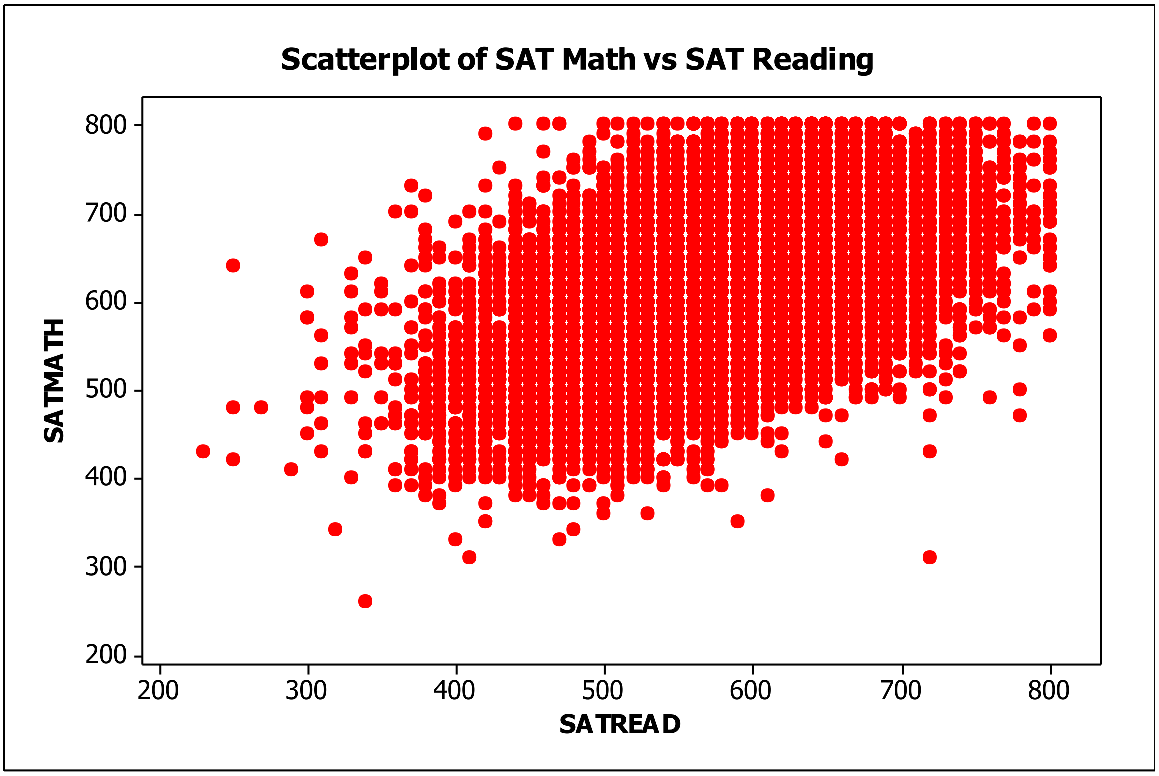


Figure 6 above is a scatterplot describing the relationship between SAT Math and SAT Reading scores. As seen in the graph, it looks like the variables have a positive linear relationship. The correlation value for this relationship is 0.496, which implies a moderate linear relationship between the two variables. This result may give rise to some minor multicollinearity issues in the model results. Both of the variables could be explaining the same variation in predicting success in the data. However, since the correlation is only moderately strong, it should not be that major a problem.

Table 5: Data Demographics for ACT

ACT	Count	Percentage
Yes	1616	44.06%
No	2052	55.94%

Table 5 shows the data distribution with respect to the created ACT variable. To repeat, the ACT variable is a dichotomous value of Yes or No depending on whether or not a student took the ACT or not. From the data above, less than half of the Fall 2006 admitted class took the ACT as well as the SAT. Since Cal Poly does not look at the ACT, it makes sense that most of the admitted students did not take the test.

Table 6: Descriptive Statistics for Quantitative Response Variables by Category

Variable	Category	Mean	StDev	Min	Max
High School GPA	Agriculture	3.53	0.39	2.45	4.45
	Science	3.77	0.33	2.97	4.50
	Engineering	3.56	0.34	2.92	4.51
	Mathematics	3.78	0.33	2.83	4.50
	Liberal Arts	3.75	0.40	2.44	4.50
	Business	3.85	0.34	2.00	4.50
SAT Math	Agriculture	563.06	73.69	310.00	780.00
	Science	610.41	67.04	440.00	750.00
	Engineering	640.06	77.55	410.00	800.00
	Mathematics	663.06	68.21	480.00	800.00
	Liberal Arts	595.79	69.98	400.00	770.00
	Business	627.72	62.67	420.00	800.00
SAT Reading	Agriculture	542.62	70.56	370.00	770.00
	Science	583.33	68.29	390.00	760.00
	Engineering	547.98	82.05	310.00	780.00
	Mathematics	586.12	73.36	410.00	800.00
	Liberal Arts	586.52	68.92	410.00	800.00
	Business	579.58	67.79	370.00	790.00

Table 6 shows the descriptive statistics for the three quantitative responses of interest by the Category of a student's background. For the High School GPA variable, the Business category has the highest average GPA. As expected, the Mathematics category has the highest average score on the SAT Math test. Also as expected, the Liberal Arts group shows the highest average SAT Reading score.

Figure 7: Boxplot of High School GPA by Category

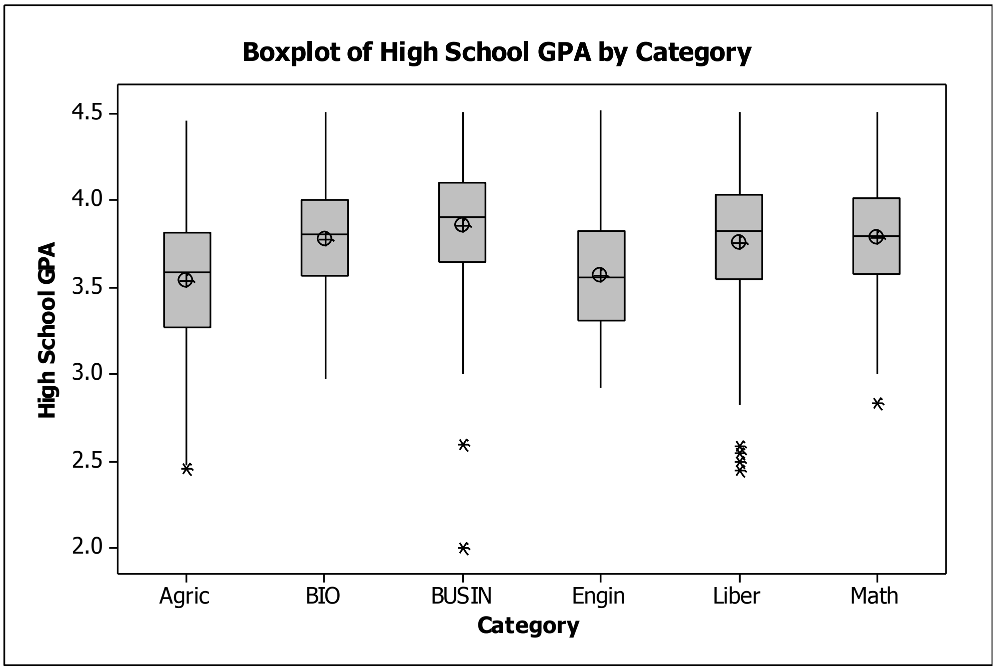


Figure 7 shows the boxplots for the distribution of High School GPA by each Category. This graph corresponds to Table 6 on the previous page. It is shown by the plots that Agriculture and Engineering have the lowest average GPA out of the six categories.

Figure 8: Boxplot of SAT Math by Category

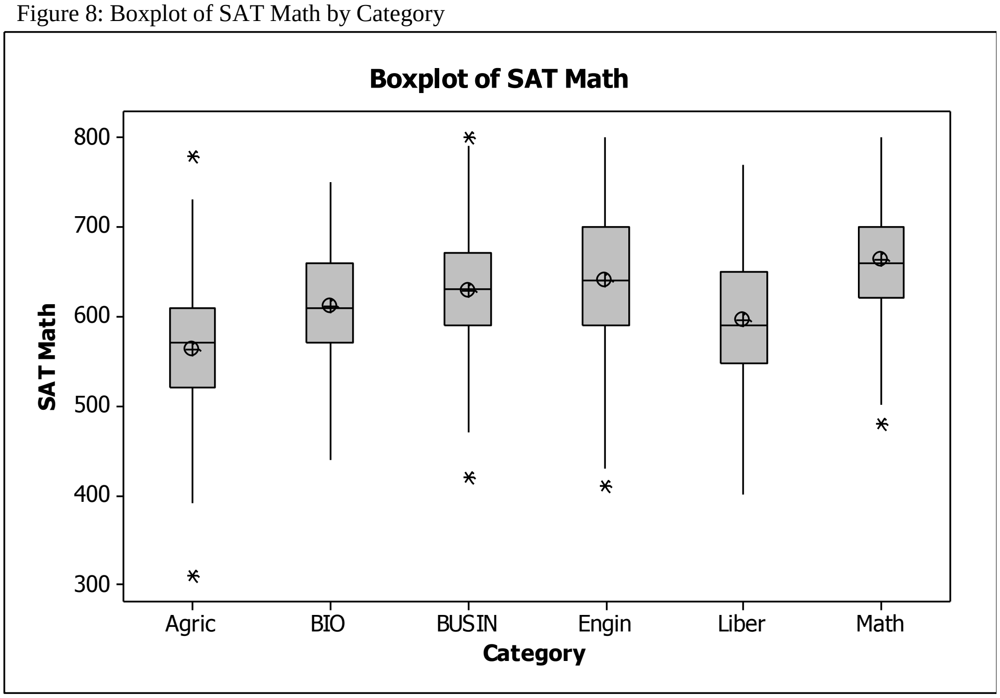


Figure 8 shows the boxplots for the distribution of SAT Math scores separated by each Category. This graph also corresponds to Table 6 on the previous page. From the distributions, Agriculture and Liberal Arts seem to have the lowest average SAT Math scores.

Figure 9: Boxplot of SAT Reading by Category

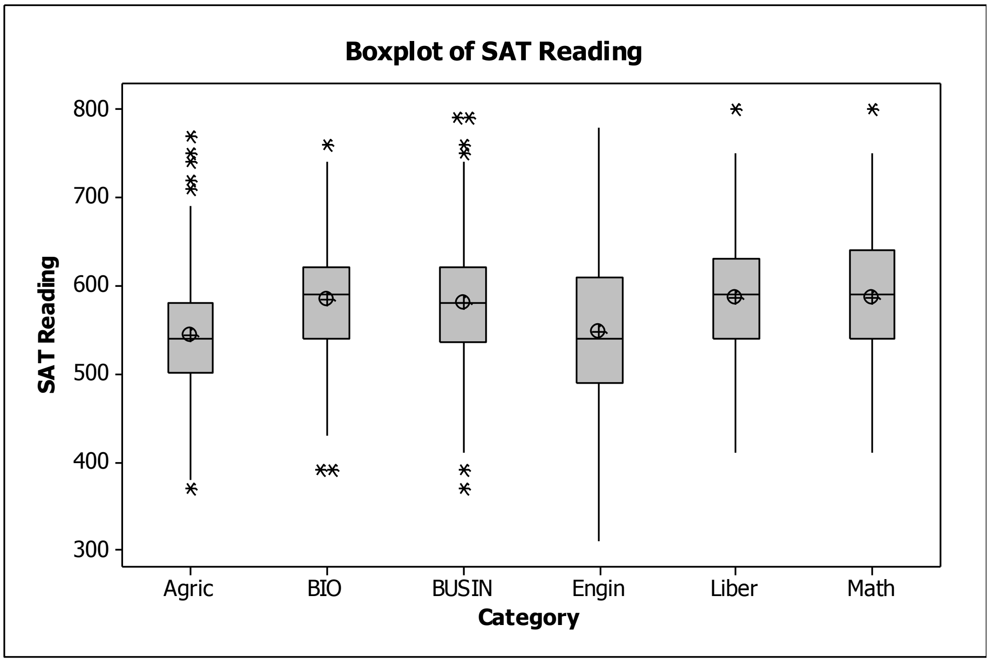


Figure 9 shows the boxplots for the distribution of SAT Reading scores based on each Category. This graph corresponds to the values in Table 6 on page 14. The distributions of scores by Category seem to be fairly even; but the Agriculture and Engineering categories seem to show the lowest average SAT Reading scores.

Based on these descriptive statistics, the analysis and results of the conducted procedures will be discussed in the next section.

VI. Analysis & Results

The results for each of the models will be presented in this section.

We begin by investigating the full model that includes the entire 2006 Fall Freshmen class. As a reminder, the response variable was the created variable Verdict and the explanatory variables were Gender, Ethnicity, ACT, SAT Math, SAT Reading, and High School GPA. The idea is to find the variables that best contribute to success of a student, while controlling for other factors. It should be noted the variables Gender and Ethnicity are included as controlling factors only. In no way does this analysis condone admitting students based on their respective Gender and Ethnicity at all. The results for the first model are presented in Table 7 below.

Table 7: Overall Model Results

Effect	**P-Value**	**Odds Ratio Estimate**	**95% Odds Ratio Interval**
Gender (Reference = Male)	0.001	1.333	(1.122 , 1.585)
Ethnicity (Reference = White)	0.001	Hispanic vs. White 0.580	(0.453 , 0.746)
HSGPA	<0.001	3.865	(3.094 , 4.828)
ACT (Reference = No)	0.124	1.134	(0.966 , 1.332)
SAT Math	0.007	1.002	(1.000 , 1.003)
SAT Reading	0.741	1	(0.999 , 1.001)

Interpretations:

Gender: Significant predictor of log-odds of success with a p-value of 0.001. With an odds ratio of 1.333, the odds of success are 33% higher for females than for males on average.

Ethnicity: Significant predictor. Since there are six levels, there are five comparisons made between each of the ethnicities and the reference group, White. The Hispanic vs. White comparison odds ratio was the only significant odds ratio which is why it's listed in the table. So, according to the data, Hispanic student's odds of success are 42% lower than a White's student. If not pairwise odds ratios are significant, then none will be listed for future models.

High School GPA: Highly Significant predictor. For every point increase in High School GPA, the odds of success are 286% higher on average.

ACT: Not a significant predictor of the log-odds success.

SAT Math: Significant predictor. For each one unit increase in SAT Math score, the odds of success are 0.2% higher on average.

SAT Reading: Not a significant predictor.

Remarks:

It does not seem to be a surprise that High School GPA and SAT Math are significant predictors of Success as a student. However, it is surprising that SAT Reading is not a significant variable. One would have the impression that a strong Reading score would affect the odds of success in similar fashion as the corresponding Math score; but it seems as if Reading does not predict success as strongly. From the overall model, we can now use these results and compare them to the results of the six individual major models.

The next six results are those from the individual major categories. Inference based on the models is deciding which factors are similar or different from the overall model.

Agriculture Model Results:

The students that are included in this model are listed in Table 1. This is true with each of the six models that will be introduced.

The Agriculture model results are listed in Table 8 below.

Table 8: Agriculture Model Results

Effect	P-Value	Odds Ratio Estimate	95% Odds Ratio Interval
Gender (Reference = Male)	0.035	2.339	(1.063 , 5.150)
Ethnicity (Reference = White)	0.591		
HSGPA	0.049	2.505	(1.001 , 6.269)
ACT (Reference = No)	0.810	0.369	(0.334 , 1.486)
SAT Math	0.007	1.008	(1.002 , 1.013)
SAT Reading	0.017	0.993	(0.988 , 0.999)

Remarks:

The significant terms in this model are listed:

- Gender
- HSGPA
- SAT Math
- SAT Reading

There are some significant differences in this model predicting log-odds for students with an Agriculture background from the overall model.

- Ethnicity is no longer significant.
- High School GPA is still significant; but does not have as strong of an association.
- SAT Reading is now a significant predictor; but according to the odds ratio, for every increase in SAT Reading score, the odds of success decrease by .007%. This result seems odd to me because one would believe that the better an individual does on a Standardized Test, the higher probability of success they will have. However, that assumption is not true in this case. Note that this results could also be the result of a Type I error due to my choice of success for each student.

Science Model Results:

The Science model results are listed in Table 9 below.

Table 9: Science Model Results

Effect	P-Value	Odds Ratio Estimate	95% Odds Ratio Interval
Gender (Reference = Male)	0.764	0.879	(0.380 , 2.033)
Ethnicity (Reference = White)	0.569		
HSGPA	0.276	2.007	(0.573 , 7.031)
ACT (Reference = No)	0.224	0.609	(0.273 , 1.354)
SAT Math	0.936	1.000	(0.993 , 1.006)
SAT Reading	0.125	1.005	(0.999 , 1.012)

Remarks:

There are no significant variables in this model. This result could be due to a lack of sample size.

Some significant differences between the Science model and the overall model include:

- Gender is no longer significant.
- Ethnicity is no longer significant.
- High School GPA does not show a significant association with the odds of success.
- SAT Math and SAT Reading are both not significant variables

Engineering Model Results:

The Engineering model results are listed in Table 10 below.

Table 10: Engineering Model Results

Effect	**P-Value**	**Odds Ratio Estimate**	**95% Odds Ratio Interval**
Gender (Reference = Male)	0.024	0.063	(1.122 , 1.585)
Ethnicity (Reference = White)	0.507		
HSGPA	0.004	6.204	(1.807 , 21.301)
ACT (Reference = No)	0.629	1.228	(0.533 , 2.828)
SAT Math	<0.001	1.012	(1.005 , 1.018)
SAT Reading	0.719	0.999	(0.994 , 1.004)

Remarks:

The significant terms in this model are listed:

- Gender
- HSGPA
- SAT Math

There is only one difference between this model and the overall model:

- Ethnicity is not a significant variable.

Mathematics Model Results:

The Mathematics model results are listed in Table 11 below.

Table 11: Mathematics Model Results

Effect	P-Value	Odds Ratio Estimate	95% Odds Ratio Interval
Gender (Reference = Male)	0.148	2.236	(0.753 , 6.643)
Ethnicity (Reference = White)	0.943		
HSGPA	0.104	3.215	(0.787 , 13.134)
ACT (Reference = No)	0.394	1.540	(0.571 , 4.156)
SAT Math	0.205	1.006	(0.999 , 1.013)
SAT Reading	0.108	1.006	(0.997 , 1.015)

Remarks:

There are no significant variables in this model. This result could be due to a lack of sample size.

Some significant differences between the Mathematics model and the overall model include:

- Gender is no longer a significant variable.
- Ethnicity is no longer a significant variable.
- High School GPA is not significant is helping predict the odds of success.
- SAT Math and SAT Reading are both not significant predictors of the odds of success.

Liberal Arts Model Results:

The Liberal Arts model results are listed in Table 12 below.

Table 12: Liberal Arts Model Results

Effect	**P-Value**	**Odds Ratio Estimate**	**95% Odds Ratio Interval**
Gender (Reference = Male)	0.689	1.179	(0.526 , 2.642)
Ethnicity (Reference = White)	0.464		
HSGPA	0.056	2.500	(0.976 , 6.405)
ACT (Reference = No)	0.144	1.763	(0.824 , 3.773)
SAT Math	0.902	1.000	(0.994 , 1.006)
SAT Reading	0.908	1.000	(0.993 , 1.006)

Remarks:

There are no significant variables in this model. The HSGPA variable is on the cusp of being significant; but with an alpha level of 0.05, it is not.

Some significant differences between the Liberal Arts model and the overall model include:

- Gender is no longer a significant variable.
- Ethnicity is no longer a significant variable.
- High School GPA is on the cusp of being significant; but is no longer significant according to the perceived alpha level of 5%.
- SAT Math is highly insignificant with a p-value of 0.902.

Business Model Results:

The Business model results are listed in Table 13 below.

Table 13: Business Model Results

Effect	P-Value	Odds Ratio Estimate	95% Odds Ratio Interval
Gender (Reference = Male)	0.061	0.565	(0.311 , 1.027)
Ethnicity (Reference = White)	0.394		
HSGPA	<0.001	5.098	(2.079 , 12.501)
ACT (Reference = No)	0.524	0.832	(0.472 , 1.465)
SAT Math	0.867	1.000	(0.995 , 1.006)
SAT Reading	0.615	1.001	(0.996 , 1.006)

Remarks:

The significant terms in this model are listed:

- HSGPA

Some significant differences between the Business model and the overall model include:

- Gender is not a significant variable.
- Ethnicity is not a significant variable.
- SAT Math is highly insignificant with a p-value of 0.867.

After seeing apparent differences in results for each of the individual major models with respect to the overall model, it is of interest to see if there are significant interaction terms. In the following model, there are seven main effect predictors as well as three interaction terms. The goal of this model is to see if there are differences in SAT Math, SAT Reading, and High School GPA by the major Category that a student is associated with.

Interaction Model Results:

The Interaction model results are listed in Table 14 below.

Table 14: Interaction Model Results

Effect	**P-Value**
Gender (Reference = Male)	0.061
Ethnicity (Reference = White)	0.394
HSGPA	<0.001
ACT (Reference = No)	0.524
SAT Math	0.867
SAT Reading	0.615
Category	0.106
HSGPA by Category	0.876
SAT Math by Category	0.062
SAT Reading by Category	0.181

According to the results above, we see that none of the interaction terms are significant. The SAT Math by Category interaction variable is on the cusp of significance with a p-value of 0.062. This result could give rise to the idea that SAT Math score predicting success has a differing effect given the background of a particular student. There is no evidence that SAT Reading and High School GPA have differing effects on the odds of success given the background of a particular student. It should also be noted that since there are interaction terms in the model, the resulting p-values for the main effect terms could be misleading. That is, we cannot tell if the Category variable itself is significant due to the presence of the interaction terms. Perhaps a future model without the interaction terms including the Category term would be of interest for determining the significance of the Category term.

VII. Summary

The goal of my statistical project was to assess the current way of admitting Cal Poly students was the most sufficient way of looking at a student's application. Based on the appropriate logistic regression procedures presented, there are some interesting results to present.

For most of the models discussed in the report, the variable High School GPA seems to have the most significant association with the log-odds of defined success. It is shown for these model results, the odds ratios were constantly higher than one when determining the difference in odds of success while holding other factors fixed. This result makes sense due to the fact that students with a higher GPA on are, on average, harder working students than those with a lower GPA. A harder working student in high school will likely be a harder working college student and therefore, on average, have a higher chance of success during their undergraduate studies. Based on this result, High School GPA should have the largest weight in determining acceptance of a student to Cal Poly. In the current MCA system, High School GPA does carry the largest weight.

With respect to the SAT scores for both Math and Reading, the Math score seems to have a much larger association with the log-odds of success than the Reading score. For a majority of the models, SAT Math was a significant predictor of success where SAT Reading was rarely significant. Based on the results, it should be considered that SAT Math and Reading scores should have separate weights, where Math is a larger weight than Reading when determining whether or not to accept a student. At the moment, the MCA criteria weights SAT Math and Reading scores equally in the Standardized Test section. There is reason to believe that different weights should be considered.

Controlling for all other variables, students that take the ACT as well as the SAT, on average, do not show a significantly higher odds of success, if one controls for SAT scores and HSGPA. At the moment, the MCA does not look at a student's ACT scores when evaluating admission. Note that I have looked at whether or not a student took the ACT, not at their particular scores. Future research in this subject could consider looking at the actual ACT scores of students. At the moment, there seems to be no association between ACT and log-odds of success.

In terms of the interaction model, none of the three interaction terms were significant. However, the SAT Math by Category variable was on the cusp of being significant. This gives reason to believe that the log-odds of success depending on a student's SAT Math score may be different based on the background of a student's study. At the moment, the MCA has equal weights for each admission criteria section for each College at Cal Poly. Although the variable was not significant, further research should be done to determine if there should be different admission criteria for each college at Cal Poly.

I believe that this project exhibits results that could contribute to the success of developing an ideal way of admitting Cal Poly students. It should be noted again that changing the definition of success as a student could drastically alter the results of the study. The admissions committee has a system that, for the most part, does a sufficient job of admitting students. Based on the results of this study, some changes could be made to improve the admissions process.

VIII. Appendix

```
      PROC IMPORT OUT= WORK.admissons
                  DATAFILE= "E:\Senior Project\Senior Project Data.xls"
                  DBMS=EXCEL REPLACE;
              RANGE="'Senior Project Data$'";
               GETNAMES=YES;
               MIXED=NO;
               SCANTEXT=YES;
              USEDATE=YES;
              SCANTIME=YES;
         RUN;

data admissions;
set admissons;

    if term ^= 2068 then delete;

    if f06tgpa ^= . then countf06=1;
     else countf06=0;
    if w07tgpa ^= . then countw07=1;
     else countw07=0;
    if s07tgpa ^= . then counts07=1;
     else counts07=0;
    if u07tgpa ^= . then countu07=1;
     else countu07=0;
    if f07tgpa ^= . then countf07=1;
     else countf07=0;
    if w08tgpa ^= . then countw08=1;
     else countw08=0;
    if s08tgpa ^= . then counts08=1;
     else counts08=0;
    if u08tgpa ^= . then countu08=1;
     else countu08=0;
    if f08tgpa ^= . then countf08=1;
     else countf08=0;
    if w09tgpa ^= . then countw09=1;
     else countw09=0;
    if s09tgpa ^= . then counts09=1;
     else counts09=0;
    if u09tgpa ^= . then countu09=1;
     else countu09=0;
    if f09tgpa ^= . then countf09=1;
     else countf09=0;

totalquarters=sum(countf06,countw07,counts07,countu07,countf07,countw08,count
s08,
                    countu08,countf08,countw09,counts09,countu09,countf09);

      if totalquarters < 7 or f08cgpa <2.30 then verdict = 'F';
     else verdict = 'S';

    if actcomp ^='' then ACT='Y';
     else ACT='N';
```

```
            if ethcode=1 then ethnicity='Hispanic';
            if ethcode=2 then ethnicity='Black';
            if ethcode=3 then ethnicity='NativeA';
            if ethcode=4 then ethnicity='Pacific';
            if ethcode=5 then ethnicity='AsianA';
            if ethcode=6 then ethnicity='MixRace';
            if ethcode=7 then ethnicity='White';
            if ethcode=8 then ethnicity='Other';
            if ethcode=9 then ethnicity='Other';

        run;

/* Subset by College */

        data code1AG;
        set admissions;
        where colgcode=1;
        run;

        data code2ARCH;
        set admissions;
        where colgcode=2;
        run;

        data code3BUS;
        set admissions;
        where colgcode=4;
        run;

        data code4ENG;
        set admissions;
        where colgcode=5;
        run;

        data code5LIB;
        set admissions;
        where colgcode=6;
        run;

        data code6MATH;
        set admissions;
        where colgcode=7;
        run;

             /*Logistic Regressions */

    proc logistic data=admissions order=data;
        class gender (param = ref ref = 'M') act (param = ref ref = 'N')
ethnicity;
        model verdict = gender ethnicity hsgpa act satread satmath;
    run;


            /*Logistic Regressions with seperation of Colleges*/
            /* College of Agriculture */

    proc logistic data=code1ag order=data;
```

```
        class gender act;
        model verdict = gender ethcode aoagroup hsgpa act satread satmath
geogreg;
    run;


           /*  College of Architecture */

    proc logistic data=code2arch order=data;
        class gender act;
        model verdict = gender ethcode aoagroup hsgpa act satread satmath
geogreg;
    run;


           /*   College of Business   */

    proc logistic data=code3bus order=data;
        class gender act;
        model verdict = gender ethcode aoagroup hsgpa act satread satmath
geogreg;
    run;


           /*     College of English   */

    proc logistic data=code4eng order=data;
        class gender act;
        model verdict = gender ethcode aoagroup hsgpa act satread satmath
geogreg;
    run;


           /*   College of Liberal Arts  */

    proc logistic data=code5lib order=data;
        class gender act;
        model verdict = gender ethcode aoagroup hsgpa act satread satmath
geogreg;
    run;


        /*   College of Science and Math  */

    proc logistic data=code6math order=data;
        class gender act;
        model verdict = gender ethcode aoagroup hsgpa act satread satmath
geogreg;
    run;



           /* Subsets for the 6 most populated Majors */

        data subAgrBus;
        set admissions;
```

```
    where majcode='AGB';
    run;

    data subArchit;
    set admissions;
    where majcode='ARCH';
    run;

    data subElecEng;
    set admissions;
    where majcode='EE';
      if majcode='EE' then category='Engin';
    run;

    data subMechEng;
    set admissions;
    where majcode='ME';
    run;

    data subBusAdmin;
    set admissions;
    where majcode='BUS';
        if majcode='BUS' then category='BUSIN';
    run;

    data subBiology;
    set admissions;
    where majcode='BIO';
        if majcode='BIO' then category='BIO';
        run;



        /*Logistic Regressions with Subsets */

proc logistic data=bigAG order=data;
     class gender act (param = ref ref = 'N') ethnicity;
     model verdict = gender ethnicity hsgpa act satread satmath;
run;

proc logistic data=bigMATH order=data;
    class gender act (param = ref ref = 'N') ethnicity;
     model verdict = gender ethnicity hsgpa act satread satmath;
run;

proc logistic data=subElecEng order=data;
    class gender act (param = ref ref = 'N') ethnicity;
     model verdict = gender ethnicity hsgpa act satread satmath;
run;

proc logistic data=bigLIB order=data;
     class gender act (param = ref ref = 'N') ethnicity;
     model verdict = gender ethnicity hsgpa act satread satmath;
run;

proc logistic data=subBusAdmin order=data;
    class gender act (param = ref ref = 'N') ethnicity;
```

```
        model verdict = gender ethnicity hsgpa act satread satmath;
    run;

    proc logistic data=subBiology order=data;
        class gender act (param = ref ref = 'N') ethnicity;
        model verdict = gender ethnicity hsgpa act satread satmath;
    run;


      /**************/


proc logistic data=interaction;
    class gender act ethnicity category;
    model verdict(descending) = category gender ethnicity hsgpa act  (param =
ref ref = 'N')  satread satmath  satmath*category;

run;
ods graphics off;
ods rtf close;



/* Ag seperation */

data bigAG; set admissions;
if majcode ^= 'REC' and majcode ^= 'DSCI'
 and majcode ^= 'NUTR'
 and majcode ^= 'ENVM'
 and majcode ^= 'EHS'
 and majcode ^= 'ERSC'
 and majcode ^= 'FDSC' then delete;
if majcode='REC' then category='Agric';
if majcode='DSCI' then category='Agric';
if majcode='NUTR' then category='Agric';
if majcode='ENVM' then category='Agric';
if majcode='EHS' then category='Agric';
if majcode='ERSC' then category='Agric';
if majcode='FDSC' then category='Agric';
run;

/* Liberal seperation */

data bigLIB; set admissions;
if majcode ^= 'ART' and majcode ^= 'COMS'
 and majcode ^= 'ENGR'
 and majcode ^= 'HIST'
 and majcode ^= 'JOUR'
 and majcode ^= 'MLL'
 and majcode ^= 'MU'
 and majcode ^= 'POLS' then delete;
if majcode='ART' then category='Liberal';
if majcode='COMS' then category='Liberal';
if majcode='ENGR' then category='Liberal';
if majcode='HIST' then category='Liberal';
if majcode='JOUR' then category='Liberal';
if majcode='MLL' then category='Liberal';
```

```
if majcode='MU' then category='Liberal';
if majcode='POLS' then category='Liberal';
run;

/* Math seperation */

data bigMATH; set admissions;
if majcode ^= 'MATH'
 and majcode ^= 'STAT'
 and majcode ^= 'PHYS' then delete;
if majcode='MATH' then category='Math';
if majcode='STAT' then category='Math';
if majcode='PHYS' then category='Math';
run;

/* INTERACTION IDEAS */

data interaction;
set bigAG subBiology subElecEng bigLIB subBusAdmin bigMATH;
run;

/* Purpose: Decide if there are different factors per college that contribute
to success */
proc logistic data = interaction;
        class gender act  (param = ref ref = 'N')  ethnicity category;
      model verdict(descending) = gender ethnicity hsgpa act   satread
satmath category
                                        satmath*category satread*category
hsgpa*category;
run;

proc means data=admissions;
var hsgpa satmath satread;
run;

proc freq data=admissions;
tables verdict;
run;

proc freq data=admissions;
tables gender*verdict;
run;
```