Bryant University

HONORS THESIS

Predictions of Success in the Actuarial Major

BY: Jessica Soojian

ADVISOR • Richard Smith

Submitted in partial fulfillment of the requirements for graduation with honors in the Bryant University Honors Program

APRIL 2016

Table of Contents

ABSTRACT

This study finds predictive factors that have a significant effect on the ending Mathematics/Actuarial GPA of Actuarial majors at Bryant University. This was done to add clarity to incoming students for what it takes to do well in the actuarial major. There were 266 subjects consisting of Bryant students that graduated between the years 2009 and 2015. The data was received in the form of final transcripts upon graduation for these students. Through manipulation of these transcripts, GPAs were calculated for Mathematics/Actuarial, English/Literary Cultural Studies, History/Politics, Economics, Science, Social Sciences, Computer Information Systems, Finance, Accounting, Management, and Marketing. These GPAs were found after 2 years (4 semesters) of classes and also after 3 years (6 semesters) of classes to see if the predictive factors differ between the two years. I also ran models comparing just the Arts and Sciences as well as just the business to see how the results differ using just those subjects rather than putting them all together. I used regression and decision trees in order to generate results. In order to use regression, I needed to impute missing values with SAS Enterprise Miner 13.2 to use all of the data. I did this using the tree surrogate method to accurately impute the data. I did not include Mathematics/Actuarial in these predictive models because of the high correlation to the final Mathematics/Actuarial GPA that I am trying to predict. After comparing the different models, English, Computer Information Systems, and Economics were the highest predictors after two years and Finance English and Science were the highest predictors after three years.

INTRODUCTION

Context

When entering college, every student, and parent, wants to make sure that they are making the right decision for a major and/or future occupation. There is a lot that goes into obtaining a college degree: money, time, effort, etc. The actuarial major is also a very competitive major. It was not always this competitive. With increased competitiveness comes increased time and effort. No one wants to get most of the way through obtaining a degree and realize that the major they chose is not what they want to be doing for the rest of their life or realize that they don't have the skills necessary to complete the major. This could be caused by many different things. One reason could be a lack of understanding of what needs to be put in to complete the major. Another reason could be that the student's interests are simply somewhere else and it takes them longer to figure that out than some others. In completing this study, I hope to identify interests and skills that are necessary to be successful in the actuarial major based on the interests and skills that make up other subject areas that may be more familiar to students than the actuarial major. In doing this, there will be huge economic benefits. If people read this study, they should take into consideration what other topics interest them and if the combination of interests adds up to those needed for the actuarial field. People who read my study will be better able to decide if the actuarial field is something that they would enjoy enough to consider doing for the rest of their life. By doing this, they will have a better understanding prior to entering college whether to rule out the actuarial major or not. This will save money and time by preventing people from investing that money and time into the actuarial major before finding out that it is not a good fit for them. This study will also have many social benefits by increasing the happiness of a person that finds out prior to investing time, money, effort, etc. that the actuarial major will be a good fit for them or if they find out that it will not be a good fit for them. This will also help the people who do decide to pursue the actuarial major and career by supporting their decision of investing time in the major and

giving them more confidence that they have made the right decision by devoting approximately four years of their life studying and many years potentially working in the actuarial field.

Lens

I am currently a senior actuarial major at Bryant University. I have not taken any actuarial exams and am set out to become a data analyst rather than an actuary. I have been an actuarial major since entering college freshman year. Although I have a strong interest in mathematics, have found that an actuarial position would not be a good fit for me. This is not to say that I or other people in my position do not have the ability to succeed in the actuarial field, but we have decided that it is not the right fit for us. This is why this project is so important to me. When I entered college, I was completely certain that the actuarial field was the perfect fit for me. At the time, there wasn't as much knowledge about the subject compared to what it is now. The actuarial field is a growing field. When I first mentioned the actuarial field four years ago, approximately only one percent of people I talked to actually knew what it was. I still get many puzzled looks from people I talk to but more people now have an idea of what it is compared to four years ago. With this growing popularity, it runs the risk of people rushing into the field for its many attractive attributes, for example, the high starting salaries and being ranked the number one job by many reputable companies for a number of years in a row. My viewpoint for this study is to help students, incoming and current, make an educated choice about what occupation to choose for their future.

PURPOSE

The purpose of this study is to give an idea to incoming freshman and current students what sort of skills are necessary to be successful in the actuarial major. I used GPAs in classes that they would be currently taking to give a different perspective that what is already out there or known about the subject. By using current GPAs to predict the success, students will have a model that they can use with the data they are collecting by taking classes. It will give students the ability to make an educated decision if this is the right major for them. The study was first done after two years because this is where most students would switch out of the major if they are going to do so. I also did the same study after three years so that you can see which factors are more important after three years compared to after two years.

LITERATURE REVIEW

Background/History

There have been numerous studies done in order to predict success. This is a very popular topic because everyone wants to maximize their success and by predicting success, one can know what their best track in life should be. This isn't always a simple task. One thing that hasn't been studied as much is the Actuarial major. Most studies are not done specifically on one topic and instead focus on predicting success on the overall college coursework. The reason there needs to be more research on a specific concentration is because some students need more guidance on a major and/or career field and this is what my study will do. A separate study involving predicting achievement in actuarial courses, in particular, could be a significant contribution, since the actuarial major requires both strong mathematical skills and other academic skills.

Review

Variables

In the research that was done on predicting academic success in all different fields, there were many different variables that were used in order to predict the success of students. These variables are categorized below and assessed accordingly. My capstone research will not be using these specific variables since they have already been researched but they will be taken into adequate consideration in my research.

Grade Point Average

In the study done by Ackerman, Kanfer, and Calderwood, they predicted academic success through the average Advanced Placement exam score. This was found to be significant and students with higher Advanced Placement scores had higher college grade point averages, took higher level classes, and graduated in less time than students with lower Advanced Placement scores. In addition to this, they found that grade point average was the best predictor of academic success. My study will be a little different because I am going to be looking at grade point averages from classes that were taken at the same time as the subject I am trying to predict success in. This study is particularly important along with other studies that find grade point average as an accurate predictor of success because this supports my decision to use grade point average as a predictive variable of success as well (Ackerman, 2013).

My study is different from all of the other subjects that prove this to be true because I am attempting to find specifically which grade point average will predict success in the actuarial major as opposed to just grade point average in general. In doing this, I am going deeper into analysis to find specifically which courses lead to better success in the actuarial major.

Though I am focusing on the actuarial major because people are not as familiar with it as other subjects, this study could also be done on other subjects as well and it would be as equally beneficial as it is for the actuarial major.

A study by Paul Bell was done in order to find what leads to success in online courses of unknown subject. Prior grade point average (academic achievement), expectancy for learning, and the interaction of those variables were significant in having an effect on the variance of academic success in online courses. This is helpful for my study because the author used prior grade point average as a variable and I plan to use a similar variable in my study. Also similarly to the Ackerman study, this study uses a prior grade point average and I am using a grade point average that was earned partially at the same time as the grade point average of what I am trying to predict success in. Since the author also found this variable to be significant, this would lead me to believe that my hypothesis may be correct in saying that the grade point average (measure of success) in all non-actuarial classes most likely has a correlation to the grade point average in actuarial classes. I will need to go into further research to determine which individual course subjects are good predictors of success in the actuarial major as opposed to all non-actuarial classes in one large group (Bell, 2007).
In the article written by Ilgan on his own study, he mentions that a study done by Beer and Beer determined that students get lower grade point averages when taking more credits and more difficult courses. This is particularly interesting because most other studies using grade point averages as predictors have tried to see what makes grade point averages go up rather than what makes them go down. This goes to show that grade point averages are universal measures of success and can be used in many different situations and used to prove many different theories in people's studies. Therefore, it will be appropriate to use this as a measure in my own study as well (Ilgan, 2013).

A study done by Juan-Claude Lemmens measured readiness as a predictor of success in higher education. High school marks, goal orientation, credits registered, learning-efficacy, gender, and parental education are some of the variables that were found significant in their study for predicting success. In this case, one of the variables they used for predicting

readiness was grade point average. By using grade point average among other variables as well, it was found that high school marks, number or registered credits, gender, and high school geographic area had an effect on academic success (Lemmens, 2001).
A study performed by Rode and Thompson discuses a study that was done to find predictive factors on SAT scores, grade point averages, and aptitude tests using working memory, processing speed, and spatial ability as variables. The results were that the Math SAT scores relied mostly on processing speed and spatial ability. In this case, grade point averages were used to measure success (Rohde, 2007).

A study done by Smith and Schumacher presents the findings of a study done to figure out the probability of an actuarial student completing the major. This data was based on SAT scores, high school grade point average, and math placement tests. This study used grade point averages to predict success in the actuarial major but this study measured success by either completing or not completing the major (Smith, 2006). My study will be using grade point average for both the predicting variable and the measuring variable of success. This means I will be able to measure how well the student is predicted to perform rather than the probability of a student completing the major.

Exam(s)

A study done by Ackerman, Kanfer, and Calderwood predicted academic success through the average Advanced Placement exam score, along with some other variables. This was found to be significant and students with higher Advanced Placement scores had higher college grade point averages, took higher level classes, and graduated in less time than students with lower AP scores. Though they found that Advanced Placement exam scores were significant, high school grade point average was a better predictor (Ackerman, 2013).

The objective nature of Advanced Placement exams makes them similar to the objective exams that students need to pass to become an actuary. There is a series of exams that you must pass to become an actuary so the results found in this study could significantly apply to

the actuarial field. But clearly this type of study has been done and we would not expect to learn much that's new in replicating it.

A study done by Abdurrahman Ilgan divides predicting factors into two categories: personality factors and contextual factors. This focused on the achievement in science courses in a college in Turkey but they found that study time, students' attitudes toward the lesson and their perception of importance were accurate predictors for how someone will perform in science courses. In determining success in these science courses, they used final examination results of the science course. This is similar to the way I plan to evaluate success in my study. The study done by Ilgan was only done on one science course (Scientific Research Method Course) and therefore it would not have made sense to use grade point average as a measure of success. Their study was done over the span of the one course so using a final grade would include grades earned early in the semester and not just the culminating point in the study. Since my research will be done on a much larger time frame, including all four years of college for each of my subjects, I can use a broader form of success with the grade point average rather than just the final examination grade. The grade point averages that I will calculate will include the final examination grades (Ilgan, 2013).
A study done by Rohde and Thompson was done to find predictive factors that have an effect on SAT scores, grade point averages, and aptitude test using working memory, processing speed, and spatial ability as variables. The results were that the Math SAT scores relied mostly on processing speed and spatial ability. This study used aptitude tests as a measure of success when working on this study. Though the predictive variables are very different than mine, the measure used to determine what is successful and what is not will be the same. This suggests that what I am choosing to use in my study as a measure of success is an accurate one and so this confirms that I am going to be successful in my study by using this measure (Rohde, 2007).

A study by Smith and Schumacher at Bryant University was done to figure out the probability of an actuarial student completing the actuarial major. This data was based on SAT scores, high school grade point average, and math placement tests. This study highly influenced the

idea for the study that I will be doing. Smith and Schumacher primarily used testing as a predicting variable. I will be using grade point averages which include any tests that were taken during each course and averaging them by the subjects. By looking at Smith and Schumacher's study, I can conclude that my predicting variable will be adequate for the project that I am looking to complete (Smith, 2006).

Qualitative and Other Factors

A study done by Bell sought to determine what leads to success in online courses of unknown subject. Prior grade point average (academic achievement), expectancy for learning, and the interaction of those variables were significant in having an effect on the variance of academic success in online courses. This study used various different predicting variables including parental level of education, computer self-efficacy, epistemological beliefs, and reasons for taking an online course. Though this study yielded interesting results, there is already a lot of research on these types of factors and their effects on success so I will not be using them in my study (Bell, 2007).

A study done by Ilgan divides predicting factors into two categories: personality factors and contextual factors. This focused on the achievement in science courses in a college in Turkey but they found that study time, students' attitudes toward the lesson and their perception of importance were accurate predictors for how someone will perform in science courses. This study took a much more psychological approach to predicting success and this could definitely be applied to my study. If the factors that I chose to determine success in the actuarial degree do not seem to fit, this would be an interesting approach to take towards finding accurate predicting factors of success (Ilgan, 2013).

A study done by Ed Jones takes a psychological approach to performance by studying self-efficacy and locus of control as predicting factors in a first year writing course. These psychological factors exist separately from certain courses so this study can be used in other topics. They found that locus of control was the highest predictor of success in a first year

writing course. Since I plan to use the grades from other classes to analyze the predicting factors, a person with internal locus of control will predictably do well in actuarial and non-actuarial classes (Jones, 2008). Locus of control transcends all areas of study so I do think this case would be useful to know about and mention in my report but this will not be my focus.

A study performed by Lemmens measured readiness as a predictor of success in higher education. High school marks, goal orientation, credits registered, learning-efficacy, gender, and parental education are some of the variables that were found significant in their study for predicting success. If a student is not ready for college when they enter, then they may receive poor grades in the beginning of their higher education career. This will lower perception of success and could then immediately lower actual success because of this. This is shown in many other studies to have an effect on success. Due to limitations in my data, this most likely cannot be tested (Lemmens, 2011).

A study by Rohde and Thompson was done to find predictive factors on SAT scores, grade point averages, and aptitude test using working memory, processing speed, and special ability as variables. The results were that the Math SAT scores relied mostly on processing speed and special ability. This study goes a step in predicting the variables that are used to predict many of the other studies done on this topic. This study uses a method similar to what I plan to do in my study and therefore this would be a useful resource for my project but there are currently many studies done on the variables that Rohde and Thompson chose to use so I do not plan on using these variables (Rohde, 2007).

The analytical study by Stankov looks into non-cognitive variables as predictors of academic achievement. This includes both social and psychological variables. Confidence was the only variable that was significant in having an effect on the variance of academic achievement. This is an important finding but confidence can come from many different things. For example, being academically successful all around could make you more confident in a

subject that you are not as strong in and increase your success in that subject. This is an important thing to consider when I am working on my study (Stankov, 2014).

Conclusions

In doing this review of literature, I have found that there is a lot of information on predicting success based on either psychological factors or high school grade point averages or other forms of testing. What I was not able to find much on is information specifically for the actuarial major. The one source by Smith and Schumacher that I found that was specifically on the actuarial major is part of what inspired my project. I also did not find a study that used other academic concentrations as a predictor of success. This is what I will be doing for my project. I have found that this type of study is different from the research that is currently out there.

As for methods, most researchers used linear or logistic regression when performing their study. I will be using linear regression because my variables (grade point averages) are continuous. Logistic regression would not work under the conditions of my study.

METHOD

The first step was to obtain the data and manipulate it to get it in the form that I needed. I received the data in the form of transcripts of Bryant actuarial graduates between the years of 2009 and 2015. I sorted through the data and had to calculate the GPAs after two years and do the same after three years. I used Excel to do this. I looked at every student and decided if they were a transfer student based on the number of courses they had and if they had taken the core classes that we are required to take or not. In total, there were 266 Bryant actuarial graduates and 10 were determined to be transfers. Therefore I was able to use 256 students in the study. The reason for this is that the study needed to be recent and therefore I used the past 7 years. Bryant only had 266 actuarial graduates in that year.

The GPAs that I used in the models were English/Literary Cultural Studies, History/Politics, Economics, Science, Social Sciences, Computer Information Systems, Finance, Accounting, Management, and Marketing. I did not include the Mathematics/Actuarial GPA for the first two or three years because this is part of the final Mathematics/Actuarial GPA. Therefore, the two are very highly correlated and will skew the results. Also, it is expected that mathematical skills would have a high effect. Since this is already known, it was not included in the model. Of these variables, English/Literary Cultural Studies, History/Politics, Economics, Science, and Social Sciences were from the College of Arts and Sciences and Computer Information Systems, Finance, Accounting, Management, and Marketing were from the College of Business.

In total, I ran six models. I ran an overall model that included all of the subject area GPAs. I then ran a model with just the subjects from the College of Arts and Sciences. I did this because as a new Bryant student, you no longer have to minor in Business Administration. For the people that I studied, along with the people in my class, we all had to major in Business Administration. Because of this, we have all taken the core Business classes such as

Finance, Accounting, Management, Marketing, and Computer Information Systems. I wanted to run a model that did not include these classes for the students who are coming into Bryant University and may not be taking these courses. I then ran a model with only subjects from the College of Business in order to have a complete study. I ran my tests for each of these models for the two year GPAs and the three year GPAs.

The tools that were used for this project were Excel, SAS Enterprise Miner 13.2, SAS 9.4, and Minitab. Excel was used only for the data manipulation of the Bryant graduates' transcripts to get the GPAs I needed. I then imported the data into Enterprise Miner 13.2. Enterprise Miner allowed me to run different models as well as impute missing data. The models I ran were decision trees, stepwise regression, and neural networks. I did not decide to include the neural networks because the data set I had was relatively small. By running a neural network, I feared of overfitting the data and decided this would not be a good model to use for this data set. The reason it may overfit the data is that a neural network uses multiple nonlinearities to make a model and therefore it is easy to overfit the model. I also did not have a large enough data set to partition the data which would have prevented overfitting. Stepwise regression was used because this type of regression adds in statistically significant variables into the model and also removes statistically insignificant variables. Forward regression only adds in statistically significant variables and backward regression only removes statistically insignificant variables and so stepwise regression is the best regression model to use.

I started by running a decision tree with the original data set I created. Imputation of missing data was necessary in order to run the regression so I did this after the decision tree. Without imputing the missing data, the program will have to throw out all the cases that contained missing data. Therefore, I used Enterprise Miner's tree surrogate method of imputation. This is a very accurate method of imputing data because it creates a decision tree with the missing variable as the target variable and imputes the data using that (Imputing Missing Values, 2016). I then ran a decision tree and a regression after the imputation. The process flow in SAS Enterprise Miner is shown in Appendix A and B. After running the models in Enterprise Miner, I did a model comparison to see which model was the best model based on average

squared error. Average squared error was used to determine the accuracy of the model because it tells you how much the predicted GPA varies from the actual GPA. I also ran regressions in both Minitab and SAS 9.4 because it gives an output in a form that is easier to interpret than SAS Enterprise Miner 13.2.

SAS 9.4 was also used to test the assumptions of regression. The assumptions tested were homoscedasticity, normality of the data, as well as independence. In the graphs and plots shown in Appendix C, you can see that each of the models do approximately fit each of the assumptions of regression. This is shown by the randomness of residuals in the top left residual plot, the points in the P-P plot following the diagonal line in the middle left plot, as well as the histogram in the bottom left appearing as a normal curve.

RESULTS

Two Year Overall

The first model I ran was the two year overall model. The decision tree is shown in Appendix D. The regular decision tree had less error than the imputed decision tree and so this is the tree I used. The first split in the decision tree was decided by their English GPA. Students with an English GPA of 3.1625 or higher were predicted to have a Mathematics/Actuarial GPA of 3.4382 and there were 152 students in this group. Students with an English GPA lower than 3.1625 were predicted to have a Mathematics/Actuarial GPA of 3.0102 and there were 104 students in this group. This makes the English GPA the best predictor of this model. The decision tree chose English as the first split because it had the highest log worth of the model. The next highest predictor was the Computer Information Systems GPA. Students with a Computer Information Systems GPA of 3.15 or higher as well as an English GPA of 3.4382 or higher were predicted to have a Mathematics/Actuarial GPA of 3.4849 and there were 138 students in this group. Students with a Computer Information Systems GPA lower than 3.15 as well as an English GPA of 3.4382 or higher were predicted to have a Mathematics/Actuarial GPA of 2.9119 and there were 18 students in this group. This made the Computer Information Systems GPA the next best predictor of the two year overall model.

The stepwise regression output is shown below in Figure A for the two year overall model. By the nature of stepwise regression, all of the variables below are significant. Also, since they are based on the same 4.0 scale, the variables with the larger coefficients have the biggest effect on the model. Therefore, in the two year overall model, Finance, English, and Accounting had the largest effect on the final Mathematics/Actuarial GPA. The R-Square of this model is 53.68% and had an average squared error of .08313.

Stepwise Selection: Step 7

Variable IMP_SS Entered: R-Square = 0.5368 and C(p) = 5.5962

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	24.08756	4.81751	57.95	<.0001
Error	250	20.78293	0.08313		
Corrected Total	255	44.87049			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	0.43994	0.17584	0.52036	6.26	0.0130
IMP_ACG	0.16716	0.03485	1.91235	23.00	<.0001
IMP_CIS	0.16374	0.04481	1.10997	13.35	0.0003
IMP_FIN	0.19727	0.04311	1.74075	20.94	<.0001
IMP_ENG	0.17409	0.04846	1.07293	12.91	0.0004
IMP_SS	0.12529	0.04472	0.65262	7.85	0.0055

Figure A- Two Year Overall Regression Output

In comparing the three models for the two year overall study, the regular decision tree had the least average squared error. The average squared error for this model was .06879. According to the regular decision tree model, English, Computer Information Systems, and Economics were the best predictors of success for the Mathematics/Actuarial final grade point average. The model comparison is shown below in Figure B.

Selected Model	Model Description	Selection Criterion: Train: Average Squared Error
Y	Decision Tree	0.06879
	Imputed Tree	0.069361
	Regression	0.081183

Figure B- Two Year Overall Model Comparison

Two Year Arts and Sciences

The next model I ran was the two year Arts and Sciences model. The decision tree is shown in Appendix E. The imputed decision tree had less error than the regular decision tree and so this is the tree I used. The first split in the decision tree was decided by their History GPA. Students with a History GPA of 3.059 or higher were predicted to have a Mathematics/Actuarial GPA of 3.4285 and there were 162 students in this group. Students with a History GPA lower than 3.059 were predicted to have a Mathematics/Actuarial GPA of 2.9815 and there were 94 students in this group. This makes the History GPA the best predictor of this model. The decision tree chose History as the first split because it had the highest log worth of the model. The next highest predictor was the English GPA. Students with an English GPA of 3.2667 or higher as well as a History GPA of 3.059 or higher were predicted to have a Mathematics/Actuarial GPA of 3.5283 and there were 104 students in this group. Students with an English GPA lower than 3.2667 as well as an English GPA of 3.059 or higher were predicted to have a Mathematics/Actuarial GPA of 3.2494 and there were 58 students in this group. This made the English GPA the next best predictor of the two year Arts and Sciences model.

The stepwise regression output is shown below in Figure C for the two year Arts and Sciences model. By the nature of stepwise regression, all of the variables below are significant. Also, since they are based on the same 4.0 scale, the variables with the larger coefficients have the biggest effect on the model. Therefore, in the two year Arts and Sciences model, Science, English, and Economics had the largest effect on the final Mathematics/Actuarial GPA. The R-Square of this model is 45.81% and had an average squared error of .09726.

Stepwise Selection: Step 5

Variable IMP_SS Entered: R-Square = 0.4581 and C(p) = 6.0000

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	20.55601	4.11120	42.27	<.0001
Error	250	24.31448	0.09726		
Corrected Total	255	44.87049			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	0.80545	0.18496	1.84434	18.96	<.0001
IMP_ECO	0.15558	0.05467	0.78761	8.10	0.0048
IMP_ENG	0.16826	0.05427	0.93483	9.61	0.0022
IMP_HIS	0.11722	0.04658	0.61588	6.33	0.0125
IMP_SS	0.11403	0.05195	0.46863	4.82	0.0291
IMP_SCI	0.17252	0.05643	0.90896	9.35	0.0025

Figure C- Two Year Arts and Sciences Regression Output

In comparing the three models for the two year Arts and Sciences study, the imputed decision tree had the least average squared error. The average squared error for this model was .084059. According to the imputed decision tree model, History, English, and Economics were the best predictors of success for the Mathematics/Actuarial final grade point average. The model comparison is shown below in Figure D.

Selected Model	Model Description	Selection Criterion: Train: Average Squared Error
Y	Imputed Tree	0.084059
	Decision Tree	0.090814
	Regression	0.096554

Figure D- Two Year Arts and Sciences Model Comparison

Two Year Business

The next model I ran was the two year Business model. The decision tree is shown in Appendix F. The regular decision tree had less error than the imputed decision tree and so this is the tree I used. The first split in the decision tree was decided by their Marketing GPA. Students with a Marketing GPA of 2.85 or higher were predicted to have a Mathematics/Actuarial GPA of 3.3488 and there were 212 students in this group. Students with a Marketing GPA lower than 2.85 were predicted to have a Mathematics/Actuarial GPA of 2.8575 and there were 44 students in this group. This makes the Marketing GPA the best predictor of this model. The decision tree chose Marketing as the first split because it had the highest log worth of the model. The next highest predictor was the Accounting GPA. Students with an Accounting GPA of 3.575 or higher as well as a Marketing GPA of 2.85 or higher were predicted to have a Mathematics/Actuarial GPA of 3.5210 and there were 97 students in this group. Students with an Accounting GPA lower than 3.575 as well as an English GPA of 2.85 or higher were predicted to have a Mathematics/Actuarial GPA of 3.2035 and there were 115 students in this group. This made the Accounting GPA the next best predictor of the two year Business model.

The stepwise regression output is shown below in Figure E for the two year Business model. By the nature of stepwise regression, all of the variables below are significant. Also, since they are based on the same 4.0 scale, the variables with the larger coefficients have the biggest effect on the model. Therefore, in the two year Business model, Finance, Accounting, and Computer Information Systems had the largest effect on the final Mathematics/Actuarial GPA. The R-Square of this model is 49.08% and had an average squared error of .09102.

Stepwise Selection: Step 4

Variable IMP_MKT Entered: R-Square = 0.4908 and C(p) = 7.6301

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	22.02390	5.50598	60.49	<.0001
Error	251	22.84659	0.09102		
Corrected Total	255	44.87049			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	0.75207	0.17378	1.70481	18.73	<.0001
IMP_ACG	0.19059	0.03640	2.49488	27.41	<.0001
IMP_CIS	0.18097	0.04702	1.34861	14.82	0.0002
IMP_FIN	0.21332	0.04619	1.94137	21.33	<.0001
IMP_MKT	0.14783	0.04175	1.14115	12.54	0.0005

Figure E- Two Year Business Regression Output

In comparing the three models for the two year Business study, the regular decision tree had the least average squared error. The average squared error for this model was .089483. According to the regular decision tree model, Marketing, Accounting, and Finance were the best predictors of success for the Mathematics/Actuarial final grade point average. The model comparison is shown below in Figure F.

Selected Model	Model Description	Selection Criterion: Train: Average Squared Error
Y	Decision Tree	0.089483
	Regression	0.093943
	Imputed Tree	0.099182

Figure F- Two Year Business Model Comparison

Three Year Overall

The next model I ran was the three year overall model. The decision tree is shown in Appendix G. The imputed decision tree had less error than the regular decision tree and so this is the tree I used. The first split in the decision tree was decided by their Finance GPA. Students with a Finance GPA of 3.2667 or higher were predicted to have a Mathematics/Actuarial GPA of 3.4553 and there were 167 students in this group. Students with a Finance GPA lower than 3.2667 were predicted to have a Mathematics/Actuarial GPA of 2.9060 and there were 89 students in this group. This makes the Finance GPA the best predictor of this model. The decision tree chose Finance as the first split because it had the highest log worth of the model. The next highest predictor was the English GPA. Students with an English GPA of 3.1417 or higher as well as a Finance GPA of 3.2667 or higher were predicted to have a Mathematics/Actuarial GPA of 3.5341 and there were 133 students in this group. Students with an English GPA lower than 3.1417 as well as a Finance GPA of 3.2667 or higher were predicted to have a Mathematics/Actuarial GPA of 2.1469 and there were 34 students in this group. This made the English GPA the next best predictor of the three year overall model.

The stepwise regression output is shown below in Figure G for the three year overall model. By the nature of stepwise regression, all of the variables below are significant. Also, since they are based on the same 4.0 scale, the variables with the larger coefficients have the biggest effect on the model. Therefore, in the three year overall model, Finance, English, and Science had the largest effect on the final Mathematics/Actuarial GPA. The R-Square of this model is 62.76% and had an average squared error of .06710.

Stepwise Selection: Step 6

Variable IMP_CIS Entered: R-Square = 0.6276 and C(p) = 7.3975

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	28.16216	4.69369	69.95	<.0001
Error	249	16.70833	0.06710		
Corrected Total	255	44.87049			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	0.31535	0.16665	0.24027	3.58	0.0596
IMP_ACG	0.13479	0.03201	1.18994	17.73	<.0001
IMP_CIS	0.09120	0.04177	0.31992	4.77	0.0299
IMP_FIN	0.21538	0.03843	2.10766	31.41	<.0001
IMP_MKT	0.09683	0.03534	0.50385	7.51	0.0066
IMP_ENG	0.18010	0.04644	1.00914	15.04	0.0001
IMP_SCI	0.14600	0.05179	0.53323	7.95	0.0052

Figure G- Three Year Overall Regression Output

In comparing the three models for the three year overall study, the imputed decision tree had the least average squared error. The average squared error for this model was .056224. According to the imputed decision tree model, Finance, English, and Science were the best predictors of success for the Mathematics/Actuarial final grade point average. The model comparison is shown below in Figure H.

Selected Model	Model Description	Selection Criterion: Train: Average Squared Error
Y	Imputed Tree	0.056224
	Decision Tree	0.057347
	Regression	0.065267

Figure H- Three Year Overall Model Comparison

Three Year Arts and Sciences

The next model I ran was the three year Arts and Sciences model. The decision tree is shown in Appendix H. The regular decision tree had less error than the imputed decision tree and so this is the tree I used. The first split in the decision tree was decided by their English GPA. Students with an English GPA of 3.2675 or higher were predicted to have a Mathematics/Actuarial GPA of 3.4786 and there were 140 students in this group. Students with an English GPA lower than 3.2675 were predicted to have a Mathematics/Actuarial GPA of 3.0057 and there were 116 students in this group. This makes the English GPA the best predictor of this model. The decision tree chose English as the first split because it had the highest log worth of the model. The next highest predictor was the Economics GPA. Students with an Economics GPA of 3.7222 or higher as well as an English GPA of 3.2675 or higher were predicted to have a Mathematics/Actuarial GPA of 3.6320 and there were 54 students in this group. Students with an Economics GPA lower than 3.7222 as well as an English GPA of 3.2675 or higher were predicted to have a Mathematics/Actuarial GPA of 3.3823 and there were 86 students in this group. This made the Economics GPA the next best predictor of the three year Arts and Sciences model.

The stepwise regression output is shown below in Figure I for the three year Arts and Sciences model. By the nature of stepwise regression, all of the variables below are significant. Also, since they are based on the same 4.0 scale, the variables with the larger coefficients have the biggest effect on the model. Therefore, in the three year Arts and Sciences model, Science, English, and Economics had the largest effect on the final Mathematics/Actuarial GPA. The R-Square of this model is 53.32% and had an average squared error of .08345.

Stepwise Selection: Step 4

Variable IMP_SS Entered: R-Square = 0.5332 and C(p) = 5.6371

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	23.92363	5.98091	71.67	<.0001
Error	251	20.94686	0.08345		
Corrected Total	255	44.87049			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	0.43327	0.17937	0.48692	5.83	0.0164
IMP_ECO	0.19472	0.04747	1.40442	16.83	<.0001
IMP_ENG	0.24350	0.05370	1.71619	20.56	<.0001
IMP_SS	0.10464	0.04836	0.39065	4.68	0.0314
IMP_SCI	0.28213	0.05362	2.31037	27.68	<.0001

Figure I- Three Year Arts and Sciences Regression Output

In comparing the three models for the three year Arts and Sciences study, the regular decision tree had the least average squared error. The average squared error for this model was .079982. According to the regular decision tree model, English, Economics, and Science were the best predictors of success for the Mathematics/Actuarial final grade point average. The model comparison is shown below in Figure J.

Selected Model	Model Description	Selection Criterion: Train: Average Squared Error
Y	Decision Tree	0.079982
	Imputed Tree	0.080032
	Regression	0.083519

Figure J- Three Year Arts and Sciences Model Comparison

Three Year Business

The next model I ran was the three year Business model. The decision tree is shown in Appendix I. The regular decision tree had less error than the imputed decision tree and so this is the tree I used. The first split in the decision tree was decided by their Finance GPA. Students with a Finance GPA of 3.2667 or higher were predicted to have a Mathematics/Actuarial GPA of 3.4545 and there were 169 students in this group. Students with a Finance GPA lower than 3.2667 were predicted to have a Mathematics/Actuarial GPA of 2.8949 and there were 87 students in this group. This makes the Finance GPA the best predictor of this model. The decision tree chose Finance as the first split because it had the highest log worth of the model. The next highest predictor was the Computer Information Systems GPA. Students with a Computer Information Systems GPA of 3.5 or higher as well as a Finance GPA of 3.2667 or higher were predicted to have a Mathematics/Actuarial GPA of 3.5151 and there were 134 students in this group. Students with a Computer Information Systems GPA lower than 3.5 as well as a Finance GPA of 3.2667 or higher were predicted to have a Mathematics/Actuarial GPA of 3.2223 and there were 35 students in this group. This made the Computer Information Systems GPA the next best predictor of the three year Business model.

The stepwise regression output is shown below for the three year Business model. By the nature of stepwise regression, all of the variables below are significant. Also, since they are based on the same 4.0 scale, the variables with the larger coefficients have the biggest effect on the model. Therefore, in the three year Business model, Finance, Marketing, and Accounting had the largest effect on the final Mathematics/Actuarial GPA. The R-Square of this model is 58.96% and had an average squared error of .07337.

Stepwise Selection: Step 4

Variable IMP_CIS Entered: R-Square = 0.5896 and C(p) = 6.1348

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	26.45349	6.61337	90.13	<.0001
Error	251	18.41700	0.07337		
Corrected Total	255	44.87049			

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	0.70061	0.14942	1.61312	21.98	<.0001
IMP_ACG	0.15585	0.03318	1.61848	22.06	<.0001
IMP_CIS	0.14128	0.04192	0.83358	11.36	0.0009
IMP_FIN	0.29299	0.03665	4.68949	63.91	<.0001
IMP_MKT	0.16336	0.03384	1.70992	23.30	<.0001

Figure K- Three Year Business Regression Output

In comparing the three models for the three year Business study, the regular decision tree had the least average squared error. The average squared error for this model was .069457. According to the regular decision tree model, Finance, Computer Information Systems, and Management were the best predictors of success for the Mathematics/Actuarial final grade point average. The model comparison is shown below in Figure L.

Selected Model	Model Description	Selection Criterion: Train: Average Squared Error
Y	Decision Tree	0.069457
	Regression	0.073222
	Imputed Tree	0.073902

Figure L- Three Year Business Model Comparison

CONCLUSION

The best predictors for the two year overall model was English, Computer Information Systems, and Economics. The best predictors for the two year Arts and Sciences model was History, English, and Economics. The best predictors for the two year Business model was Marketing, Accounting, and Finance.

The best predictors for the three year overall model was Finance, English, and Science. The best predictors for the three year Arts and Sciences model was English, Economics, and Science. The best predictors for the three year Business model was Finance, Computer Information Systems, and Management.

It is important to note that when looking at the decision trees, it is not saying that you have to excel in the predicting subjects in order to do well in your Mathematics/Actuarial GPA. It is saying that you must have reasonable skills in those areas to excel. For Example, the English GPA cutoff for the two year overall model is at an English/Literary Cultural Studies GPA of 3.1625. Therefore, you only roughly need a B average to be put in the group with the higher overall Mathematics/Actuarial GPA. This just shows that you need those types of skills to translate into the actuarial work, but not necessarily being the top of the class in those subject areas.

It is also important to note that the three year models had less error than the two year models. This is most likely because there was more data to collect after three years compared to after two years. The extra data that was able to be collected allowed for a more accurate model.

It is clear that the difference in predictors between two years of data and three years of data is prominent. This may be because the curriculum in the courses may correlate more with the higher level courses rather than the lower level courses. Finance has a lot of overlap between courses in the actuarial major once you get to the higher level courses. This is why Finance becomes the best predictor after three years but is not seen after two in the overall model.

In addition to this, Arts and Sciences courses showed up as better predictors more often overall compared to the business courses. English, Economics, and Science show up in the overall models as top three predictors from the Arts and Sciences but only Finance and Computer Information Systems showed up from the Business side. Further research would have to be done to conclude why Arts and Sciences courses seem to be better predictors of success in the Actuarial Major.

IMPLICATIONS AND FUTURE STEPS

The first major implication is that this study is major specific. The models created in this study can only be used to predict a student's Mathematics/Actuarial GPA. This type of study should definitely be done for all other majors, minors, and concentrations because it leads to some very insightful results.

This study is also based purely on numbers. Because of confidentiality concerns, I was not able to have the names of the people in the study and therefore was not able to talk to them. Therefore, I could not gather any other data besides their grades in their college courses. This study could be augmented if someone were able to ask these graduated students questions along with their grades. Though this may make the study better, this study is still valid in predicting success in ones' Mathematics/Actuarial GPA upon graduation.

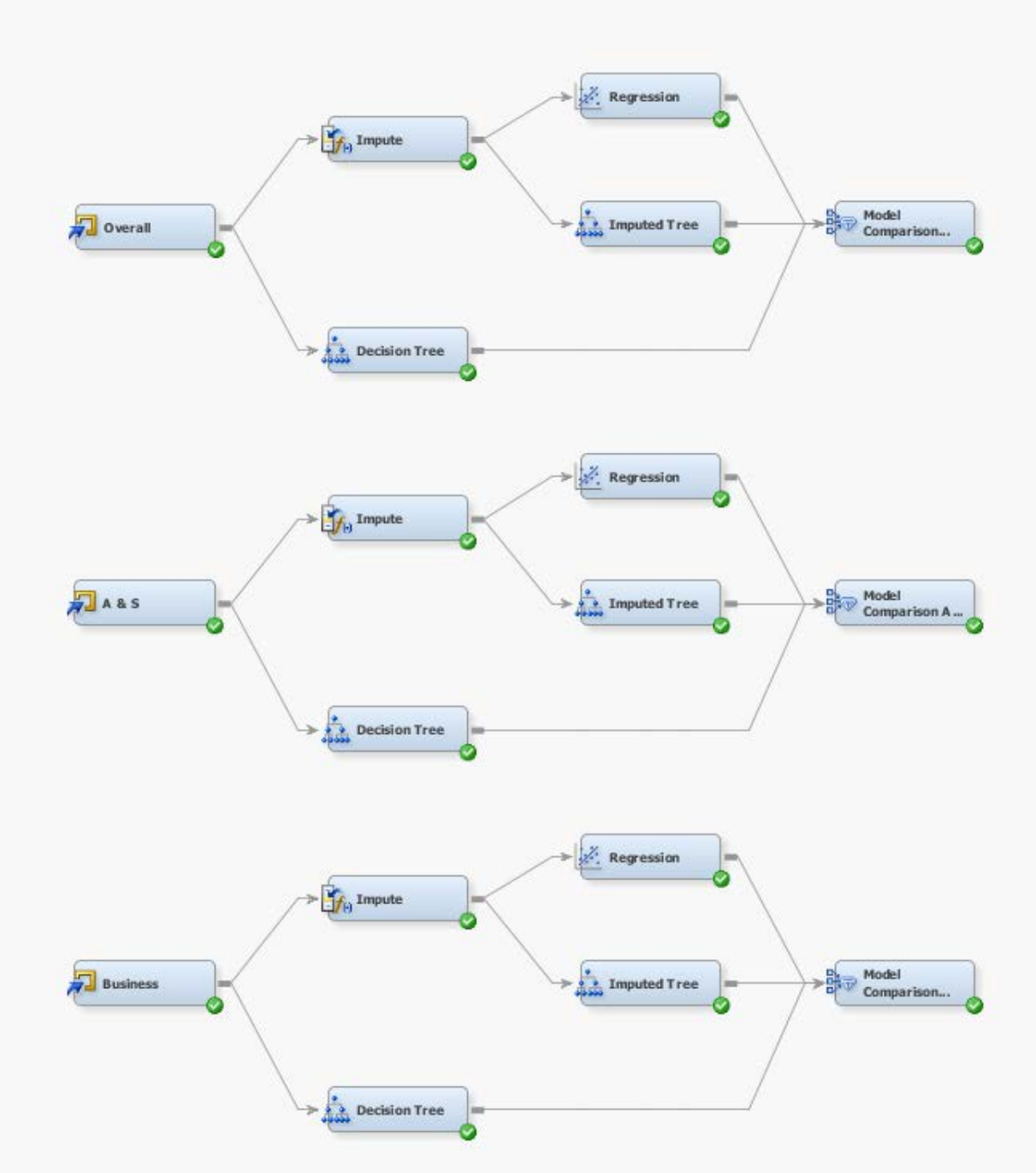
In addition to this, the GPAs were calculated after two and after three years of courses. This may be a little late in someone's college career to be making any sort of decision regarding a change in major, especially after three years, if they do not want to go to school longer than four years. That being said, if the study was done after one year, there would be too much missing data to justify the study. This is why I ran the test after two years and after three years. This still gave some valuable information and it allowed my study to contain more collected data rather than imputed data.

Because of what is said above, I would do this study at a bigger school. There is more data available at a bigger school and would allow for more freedom in the study. I only used 7 years' worth of data because I wanted the study to be more likely to have forecast ability. This significantly restricted the amount of students' transcripts that I could use. By using a bigger school, there should be enough data to run the tests after one year. I also would be able to partition the data which would prevent the model from overfitting. I did not partition the data that I had because it was a relatively small data set with only 256 Bryant actuarial

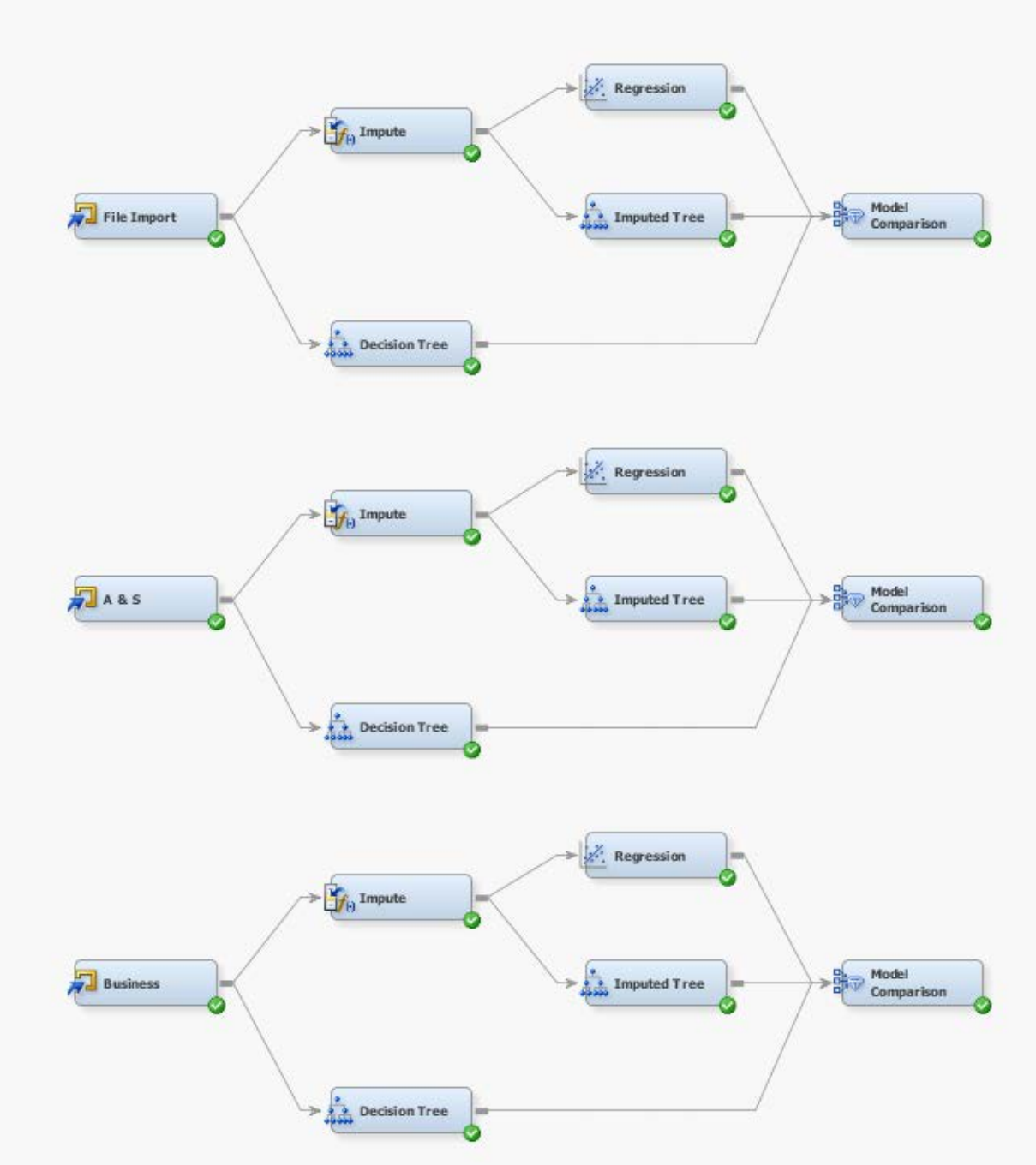
graduates. Therefore, there is a chance that the models that I created may have overfitted the data. By using a bigger school, I would be able to keep the data recent, run the data after one year of GPAs, partition the data, and create a model that would significantly reduce the chance of overfitting the data.

APPENDICES

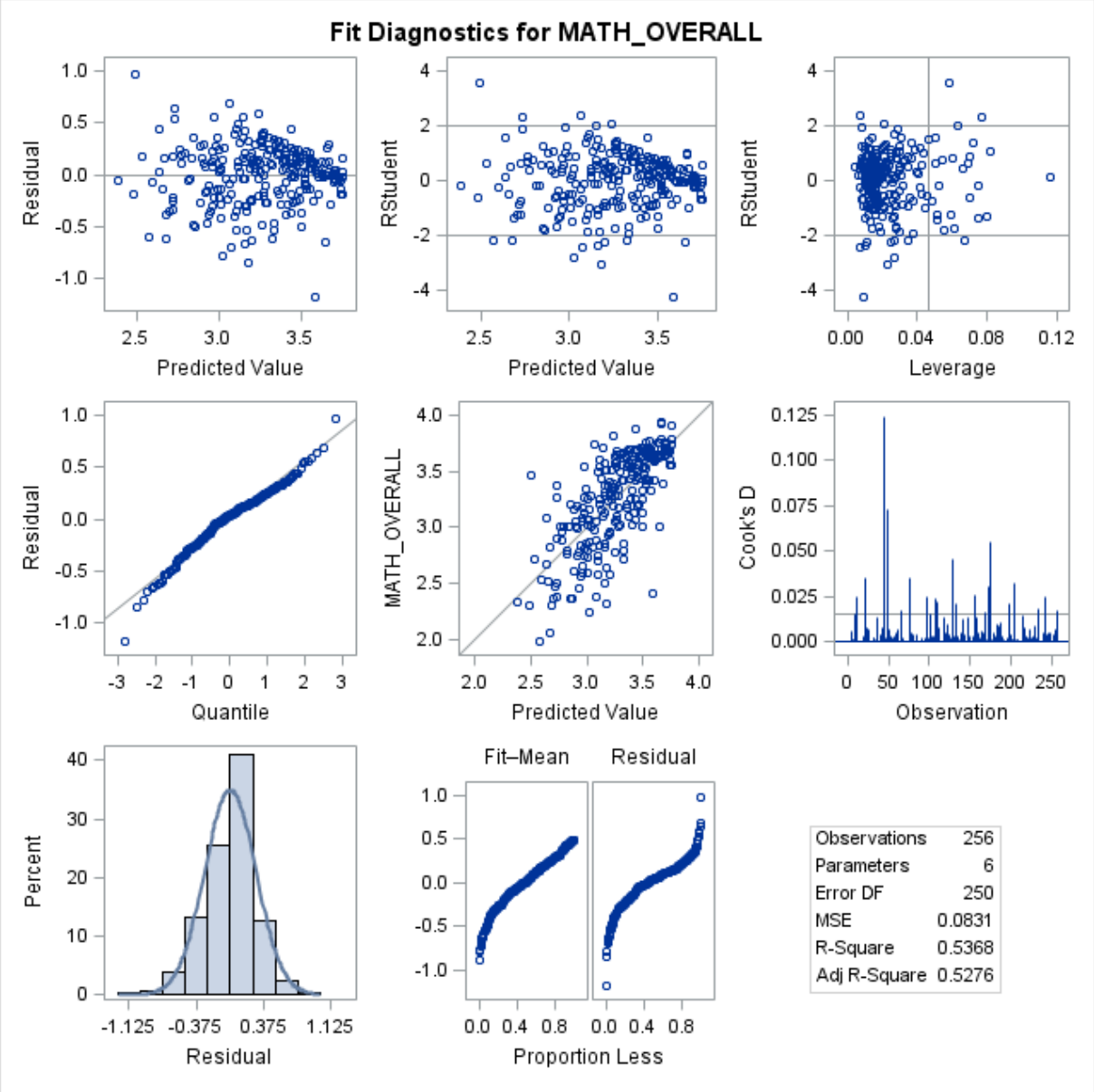
Appendix A- Enterprise Miner Process Flow Two Year



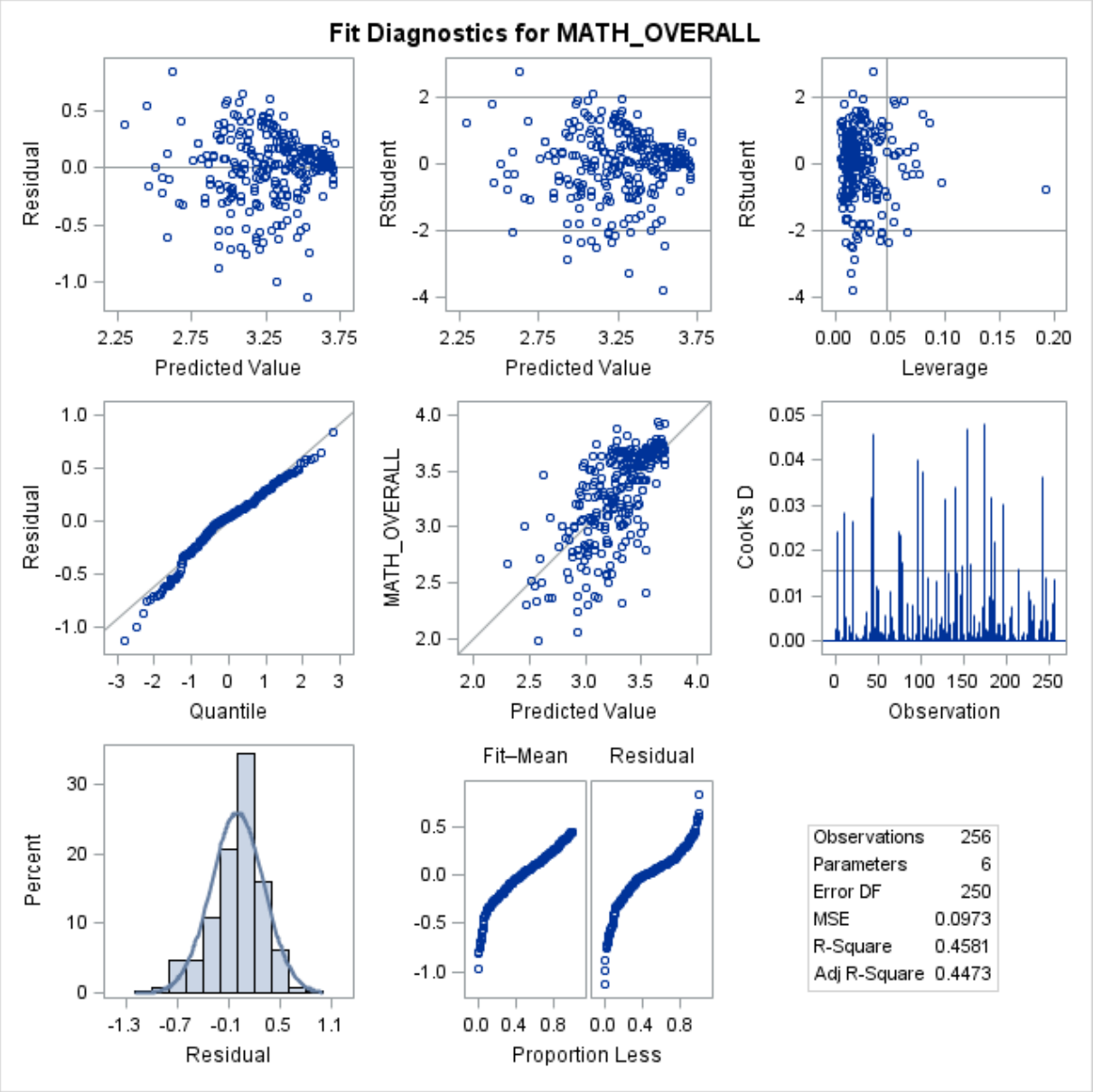
Appendix B- Enterprise Miner Process Flow Three Year



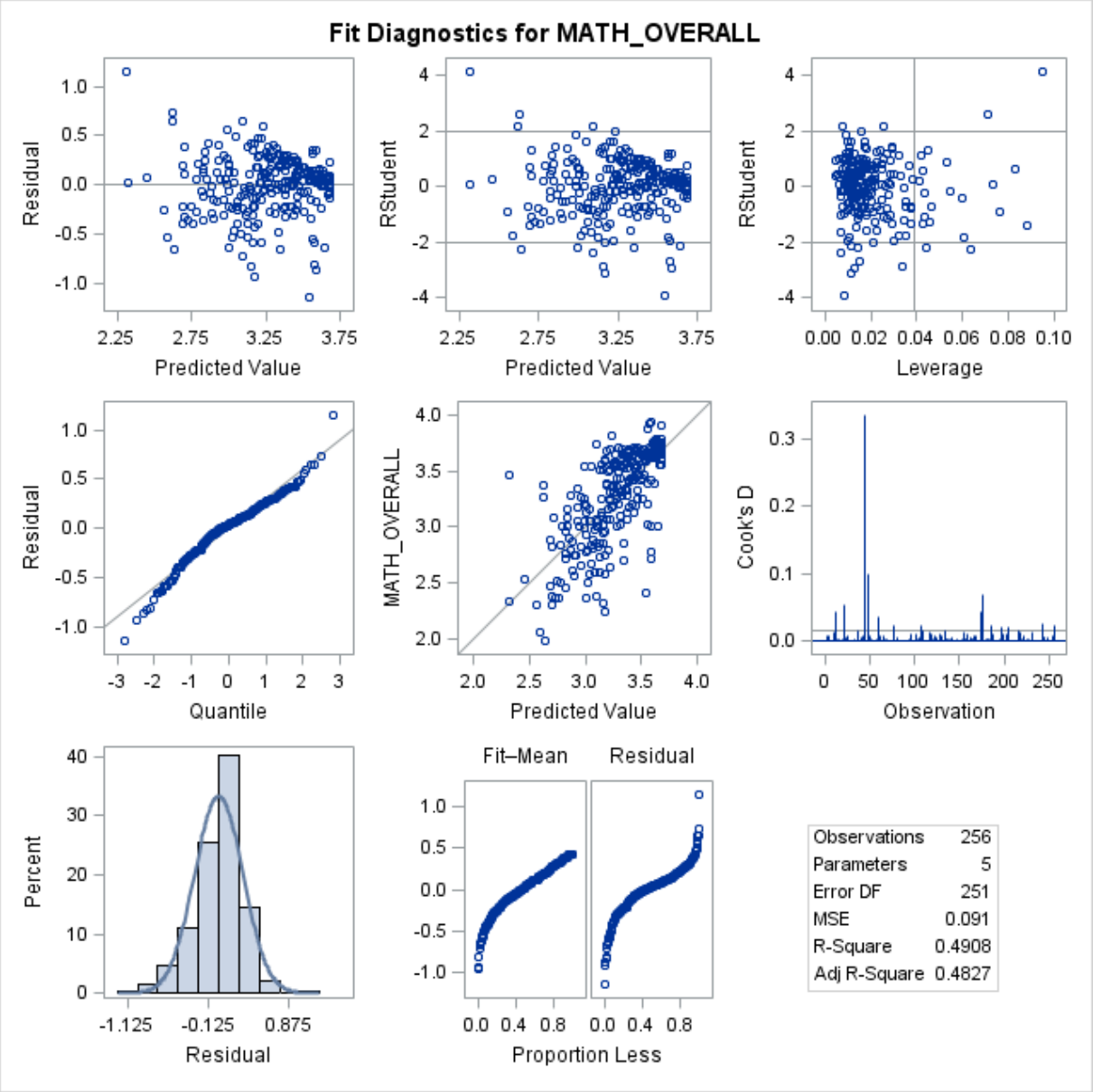
Appendix C- Test of Assumptions



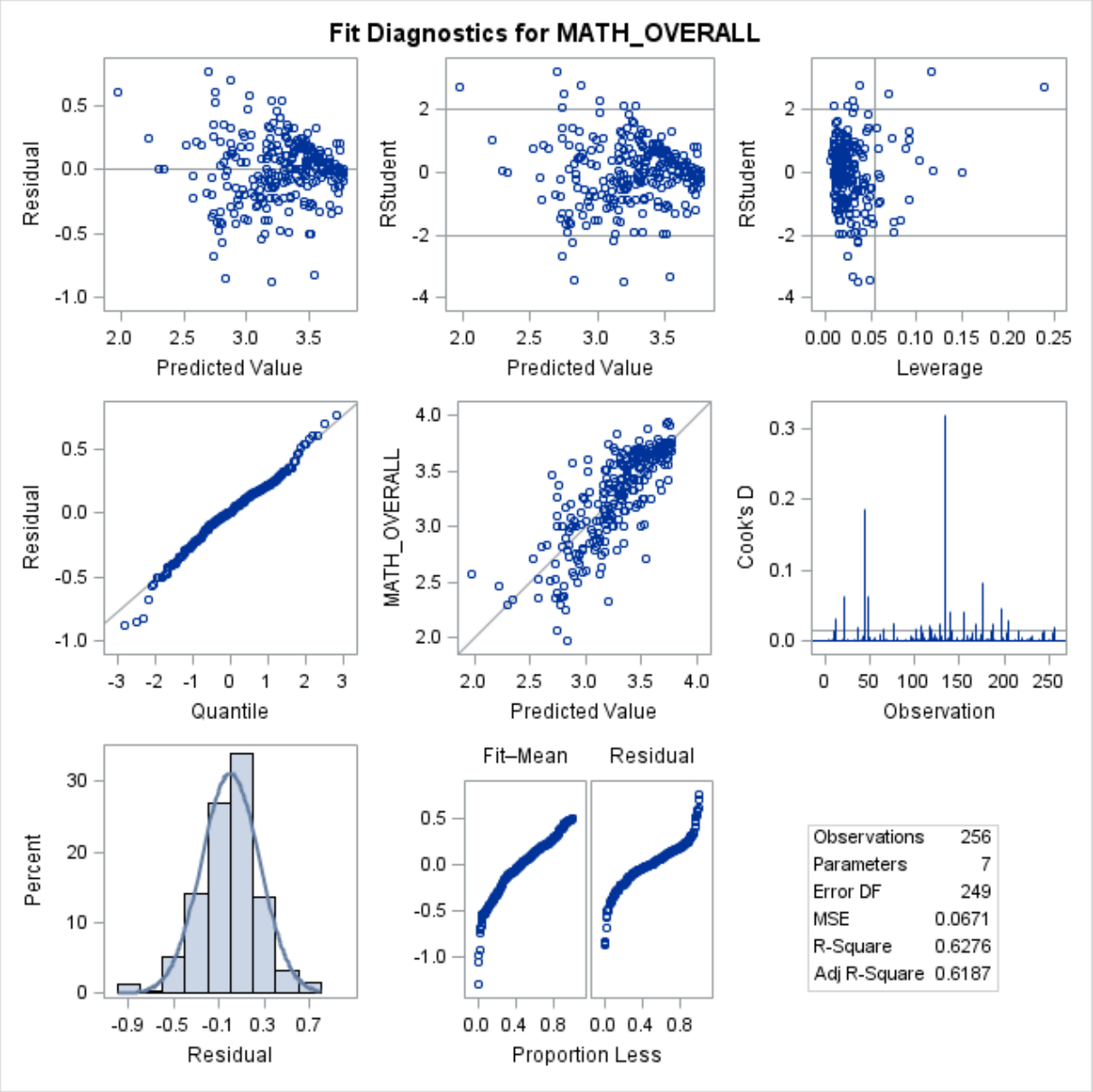
Two Year Overall Fit Diagnostics



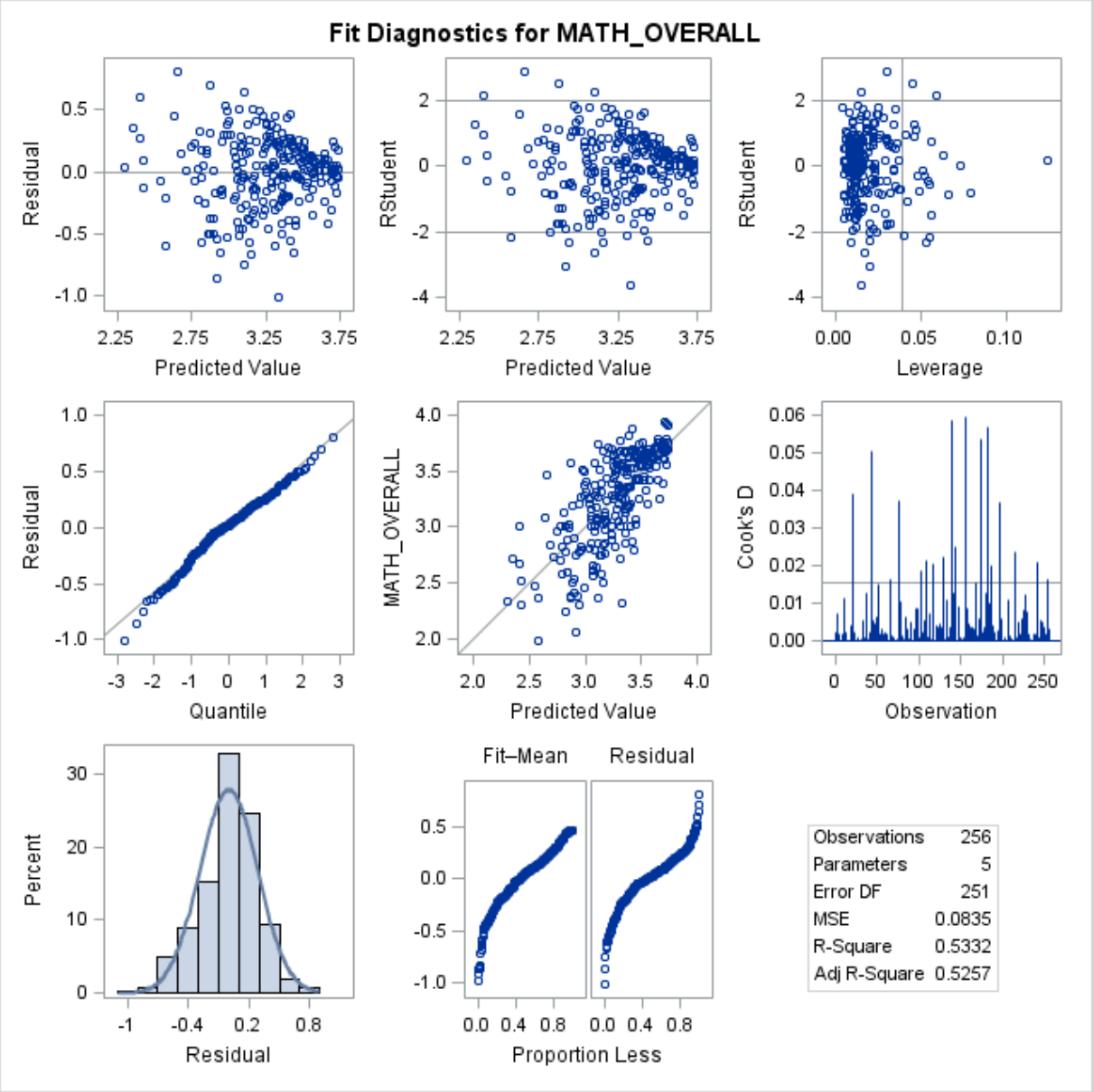
Two Year Arts and Sciences Fit Diagnostics



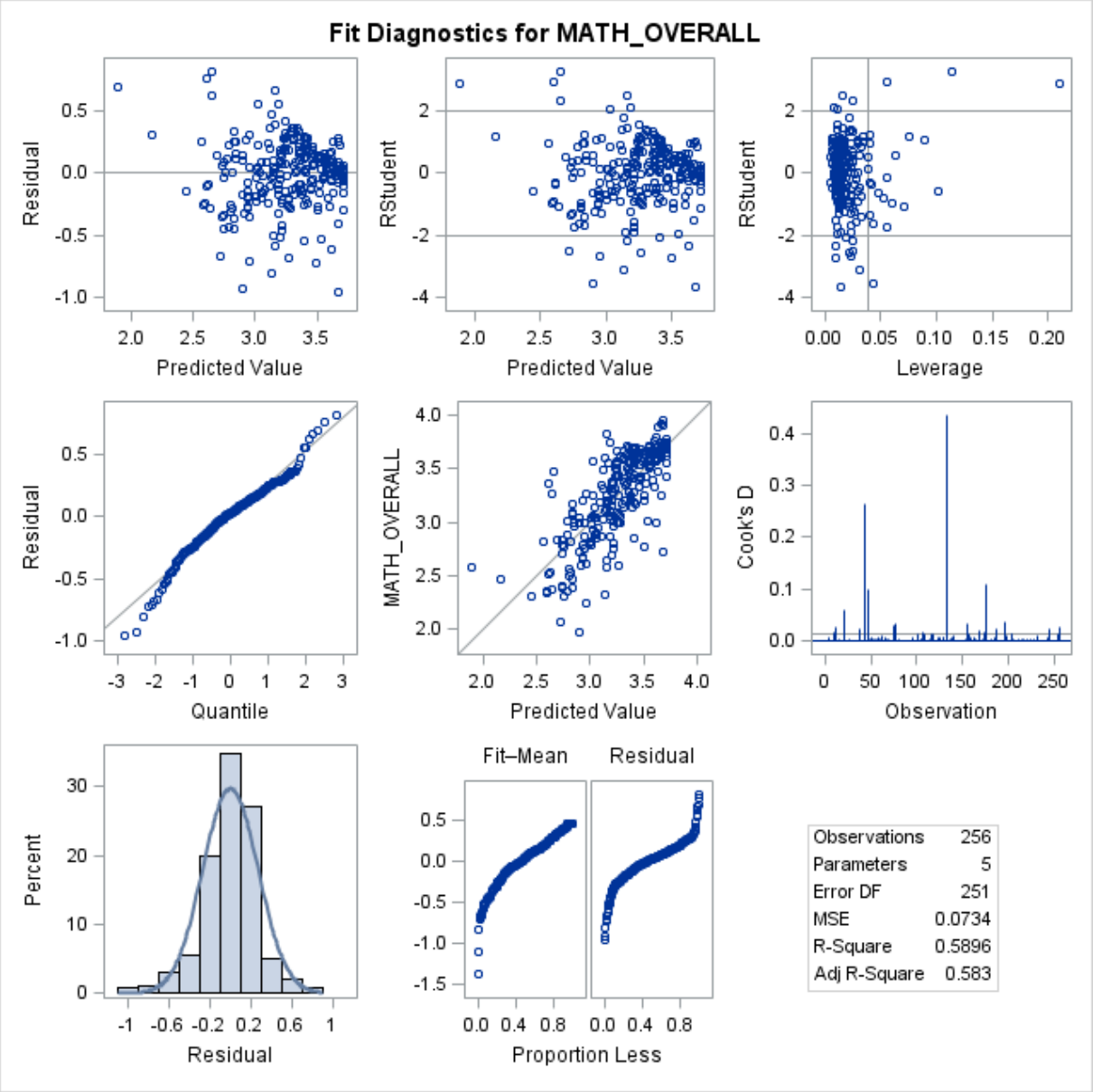
Two Year Business Fit Diagnostics



Three Year Overall Fit Diagnostics

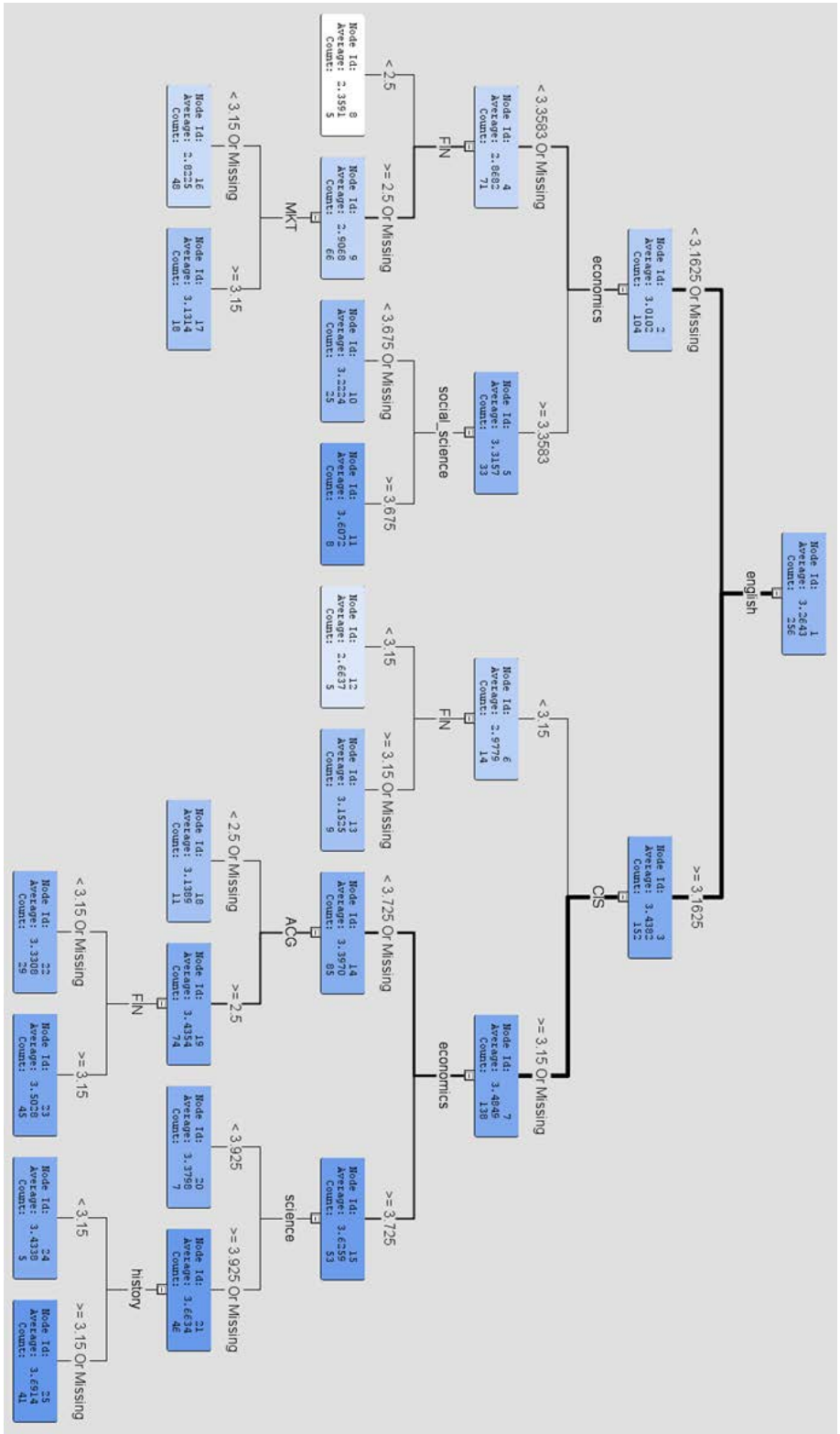


Three Year Arts and Sciences Fit Diagnostics

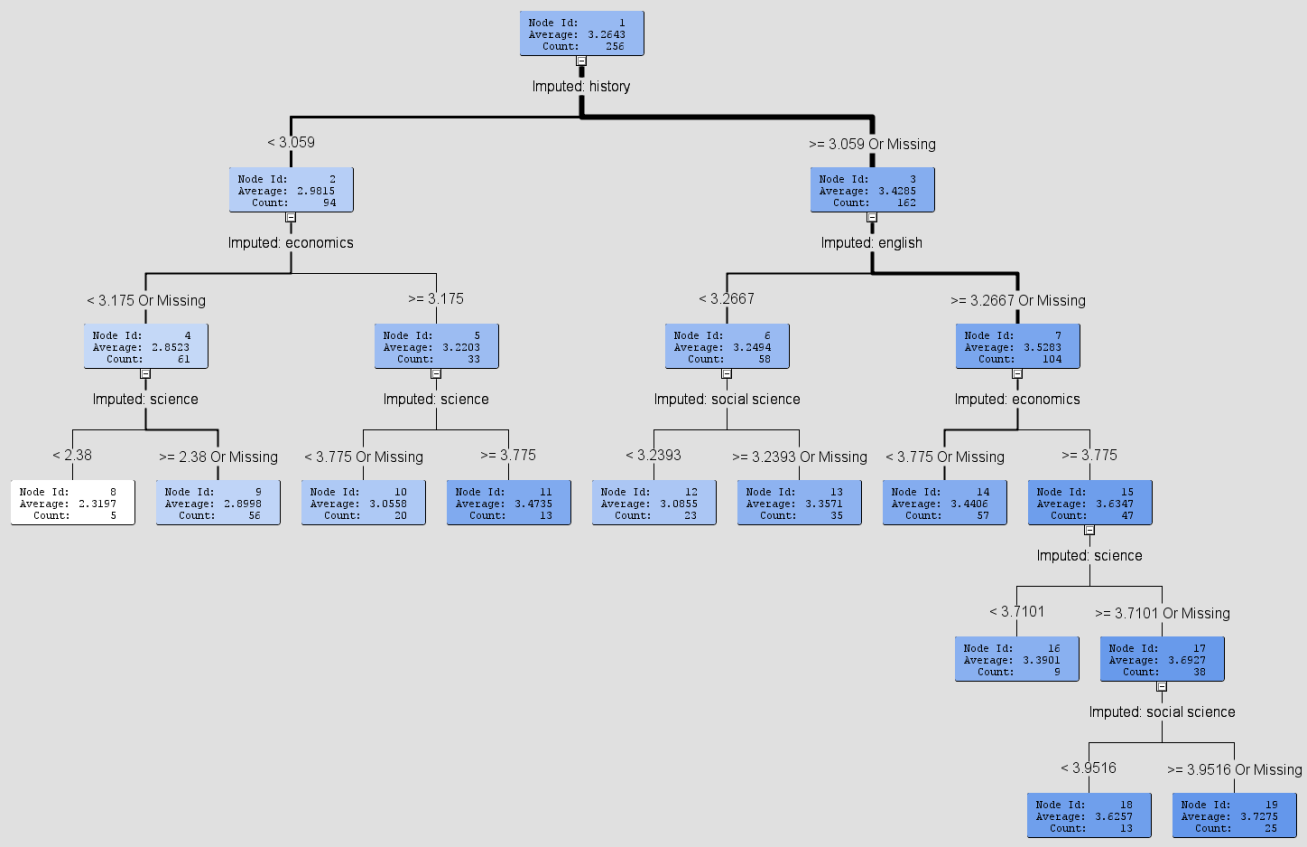


Three Year Business Fit Diagnostics

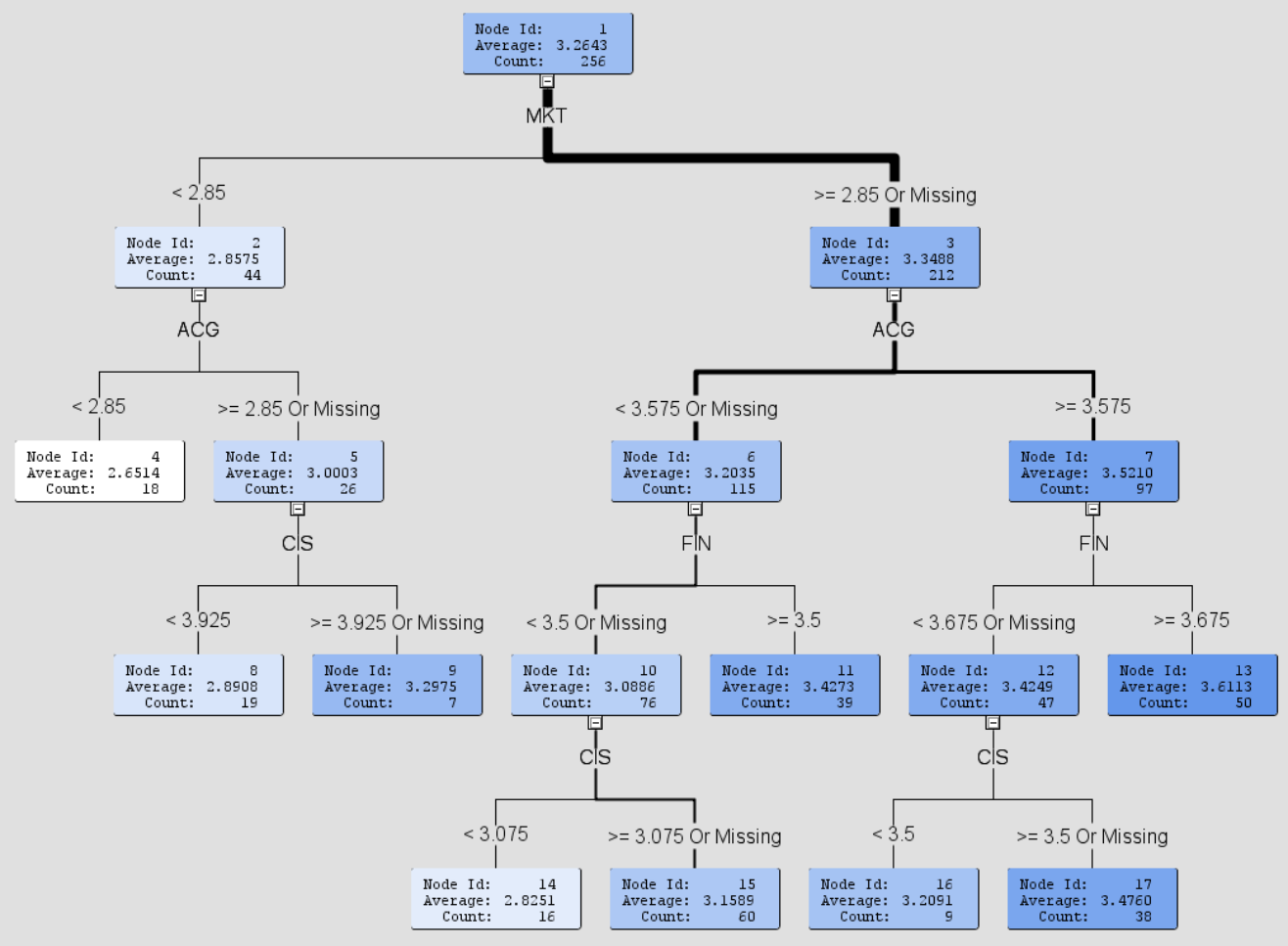
Appendix D- Two Year Overall Decision Tree



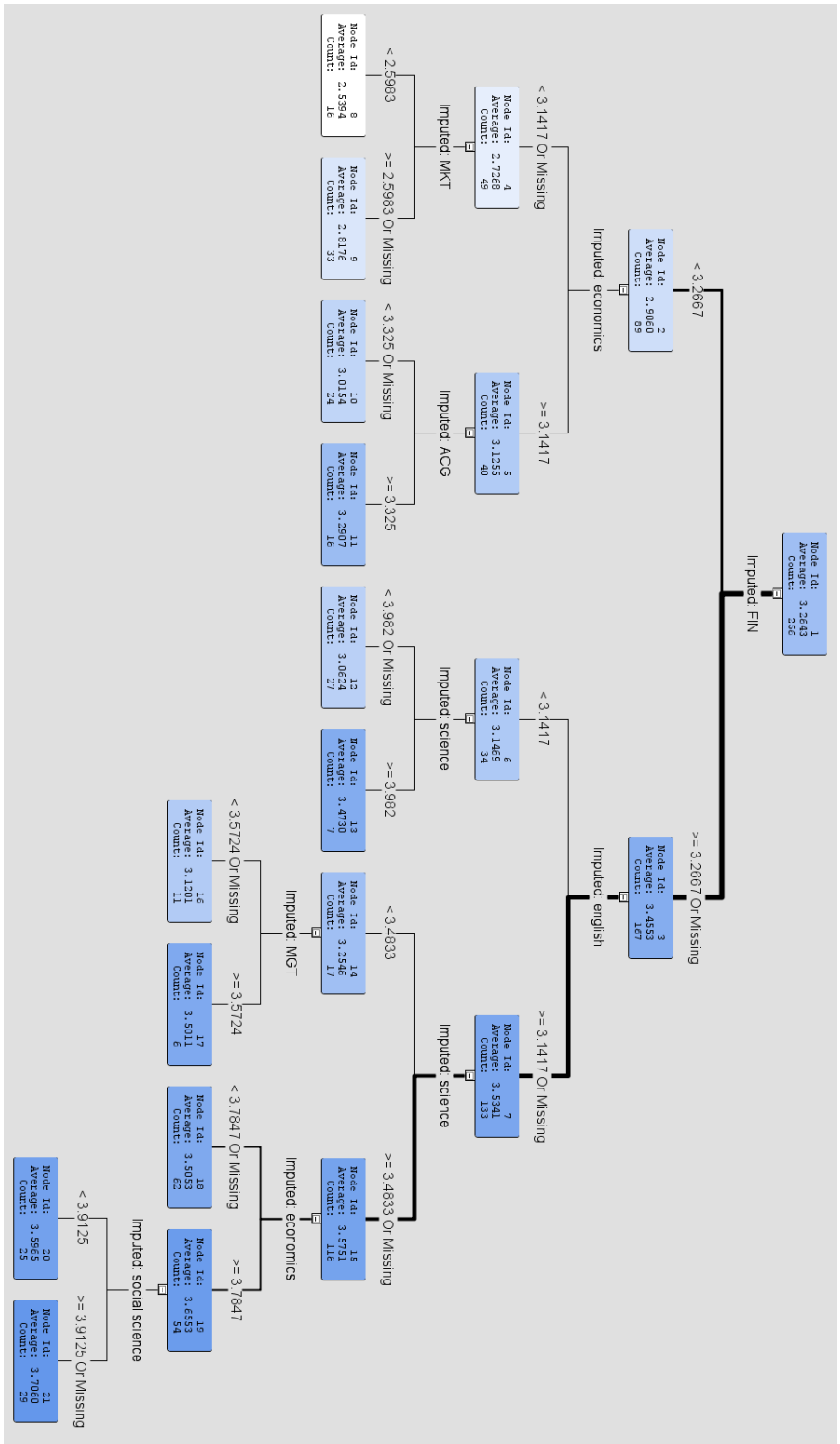
Appendix E- Two Year Arts and Sciences Decision Tree



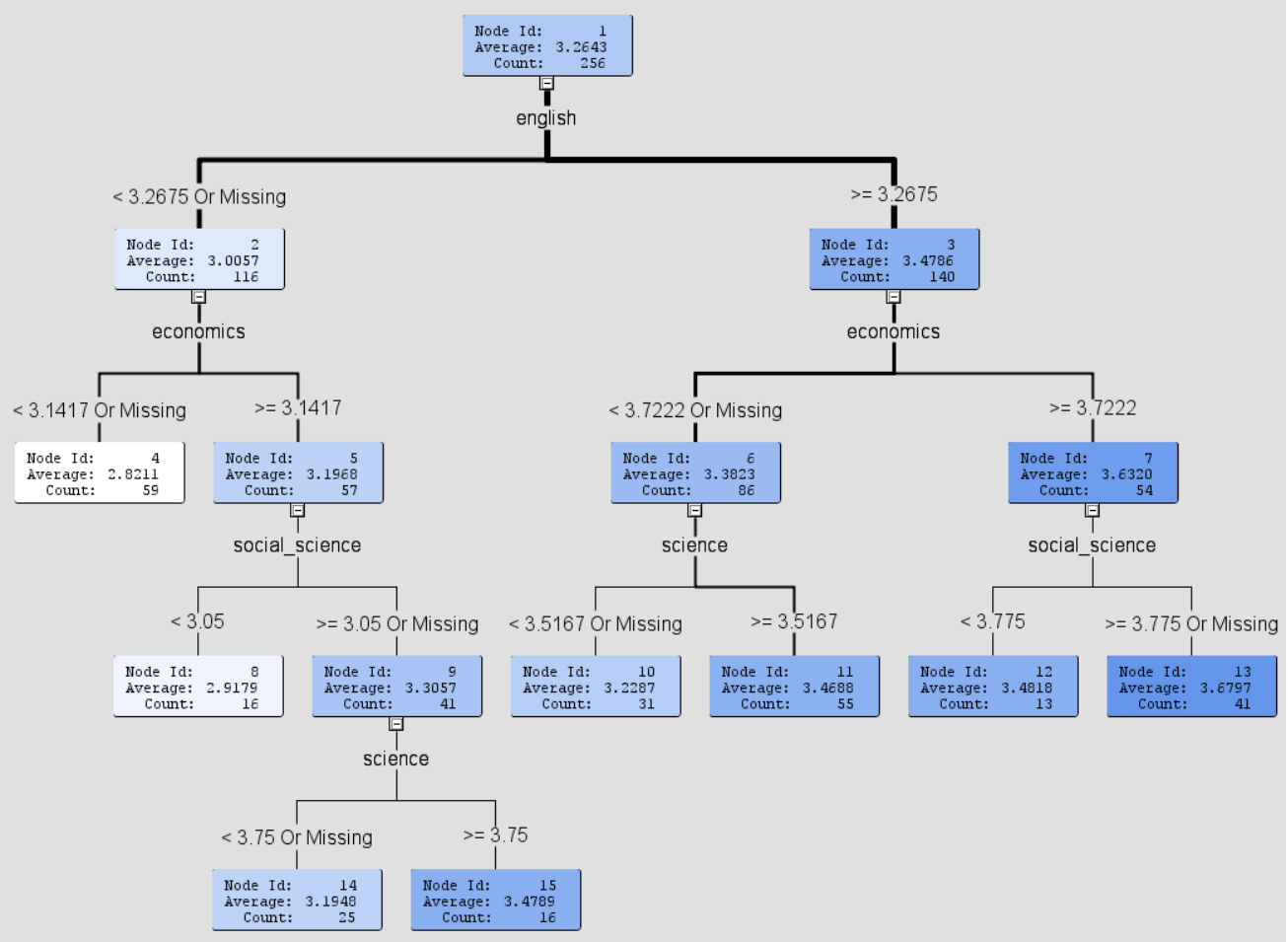
Appendix F- Two Year Business Decision Tree



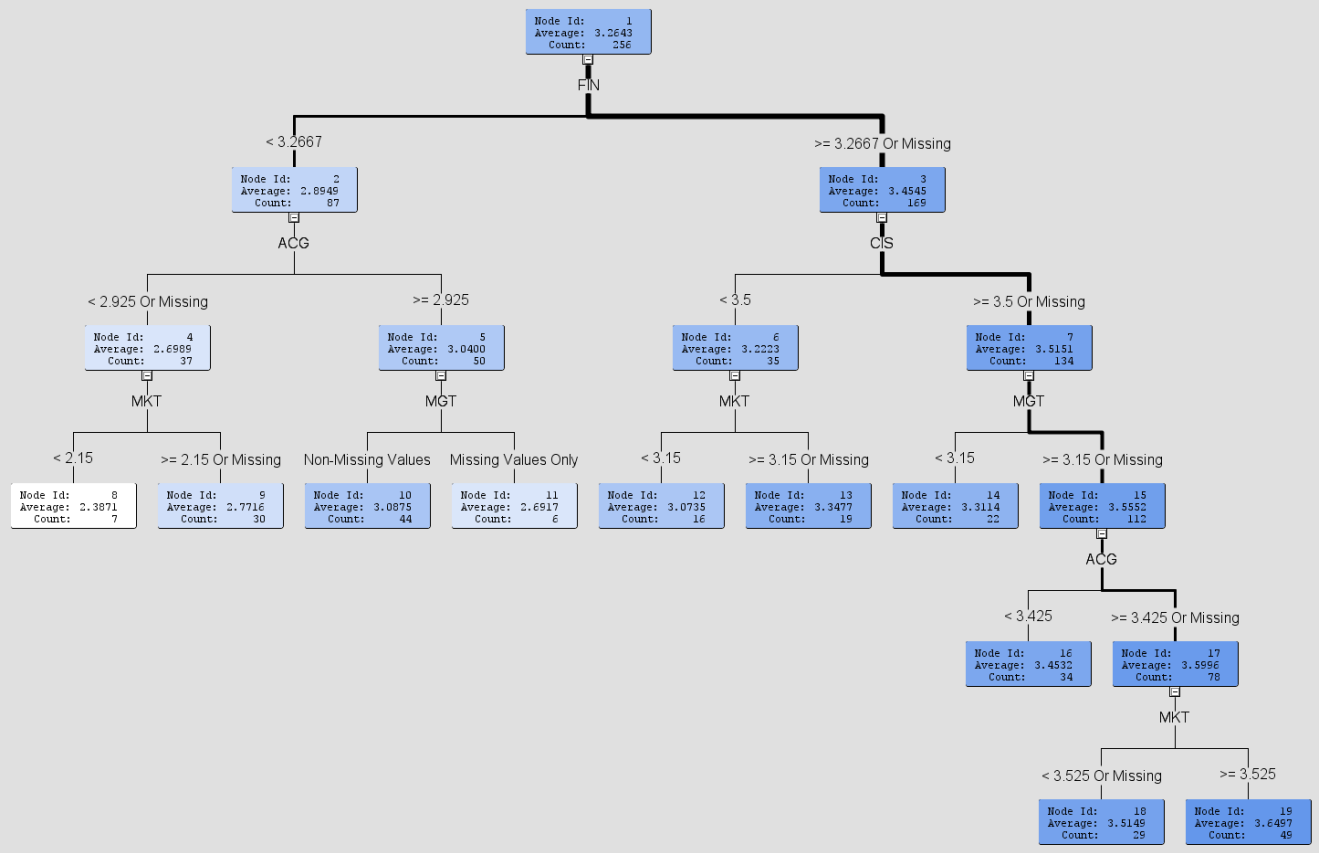
Appendix G- Three Year Overall Decision Tree



Appendix H- Three Year Arts and Sciences Decision Tree



Appendix I- Three Year Business Decision Tree



REFERENCES

Ackerman, P. L., Kanfer, R., & Calderwood, C. (2013). High School Advanced Placement and Student Performance in College: STEM Majors, Non-STEM Majors, and Gender Differences. Teachers College Record, 115(10), 1-43.

Bell, P. D. (2007). Predictors of College Student Achievement in Undergraduate Asynchronous Web-Based Courses. Education, 127(4), 523-533.

DeBerard, M. S., Spielmans, G. I., & Julka, D. C. (2004). Predictors of Academic Achievement and Retention Among College Freshmen: A Longitudinal Study. College Student Journal, 38(1), 66-80.

Gorges, J., & Göke, T. (2015). How do I know what I can do? Anticipating expectancy of success regarding novel academic tasks. British Journal Of Educational Psychology, 85(1), 75-90. doi:10.1111/bjep.12064

Hailikari, T., Nevgi, A., & Komulainen, E. (2008). Academic self-beliefs and prior knowledge as predictors of student achievement in Mathematics: a structural model. Educational Psychology, 28(1), 59-71. doi:10.1080/01443410701413753

Ilgan, A. (2013). Predicting College Student Achievement in Science Courses. Journal Of Baltic Science Education, 12(3), 322-336.

Imputing Missing Values. (2016, April 9). Retrieved from Getting Started with SAS(R) Enterprise Miner(TM) 6.1: http://support.sas.com/documentation/cdl/en/emgsj/61207/HTML/default/viewer.htm#p067358imf26jtn1ntxi6gon3lrg.htm

Jones, E. (2008). Predicting performance in first-semester college basic writers: Revisiting the role of self-beliefs. Contemporary Educational Psychology, 33(2), 209-238. doi:10.1016/j.cedpsych.2006.11.001

Lemmens, J., du Plessis, G. I., & Maree, D. F. (2011). Measuring Readiness and Success at a Higher Education Institution. Journal Of Psychology In Africa (Elliott & Fitzpatrick, Inc.), 21(4), 615-621.

Rohde, T. E., & Thompson, L. A. (2007). Predicting Academic Achievement with Cognitive Ability. Intelligence, 35(1), 83-92.

Smith, R. M., & Schumacher, P. (2006). Academic Attributes of College Freshmen That Lead to Success in Actuarial Studies in a Business College. *Journal of Education for Business, 81*(5), 256–260.

Soria, K. M., Fransen, J., & Nackerud, S. (2014). Stacks, Serials, Search Engines, and Students' Success: First-Year Undergraduate Students' Library Use, Academic Achievement, and Retention. Journal Of Academic Librarianship, 40(1), 84-91. doi:10.1016/j.acalib.2013.12.002

Stankov, L., Morony, S., & Lee, Y. P. (2014). Confidence: the best non-cognitive predictor of academic achievement?. Educational Psychology, 34(1), 9-28. doi:10.1080/01443410.2013.814194

Torenbeek, M., Jansen, E., & Hofman, W. (2011). Predicting first-year achievement by pedagogy and skill development in the first weeks at university. Teaching In Higher Education, 16(6), 655-668. doi:10.1080/13562517.2011.560378

Torenbeek, M., Jansen, E., & Suhre, C. (2013). Predicting undergraduates' academic achievement: the role of the curriculum, time investment and self-regulated learning. Studies in Higher Education, 38(9), 1393-1406. doi:10.1080/03075079.2011.640996

Vivo, J., & Franco, M. (2008). How does one assess the accuracy of academic success predictors? ROC analysis applied to university entrance factors. International Journal Of Mathematical Education In Science & Technology, 39(3), 325-340. doi:10.1080/00207390701691566