Predicting the Completion of a Preliminary Year of College Based on the Student's Attributes

McCord Murray

April 11, 2022

Abstract

The goal of this experiment is to determine if a certain combination of attributes was predictive of a student passing a preliminary year at UMass Dartmouth. To test the hypothesis that a certain combination of attributes was predictive of a student passing, we will first pick some variables that we believe make the most sense in being closely related to whether or not a student passes the preliminary year. Then, a logistic model and multinomial logistic model will be constructed and plotted to measure how well the attributes predict passing. It was found that attending workshops, advisor meetings, and peer mentor meetings during the semester improved the odds of passing, while the psychological examinations given before the semester did not predict a student passing as well as it was designed to.

Background and Significance

When colleges go through a student's application they review many different attributes of the student to determine whether or not they are a good fit for the school. However, different attributes may matter more or less depending on the reason the student is applying. Most students look to pursue a degree, in which case the college may be looking for applicants with high SAT/ACT scores and a high GPA, whereas applicants who are looking to be a full time athlete have different priorities, so their SAT/ACT scores and GPA aren't as important. Ultimately, the college is looking to admit applicants that they believe will both complete their degree and make a positive impact on the college. What we would like to know is which attributes are the best indicators of a student's success? A study was conducted at the University of Massachusetts - Dartmouth for some students who were accepted into a preliminary year. The data from this study includes, in addition to GPA, SAT scores, and whether or not they passed, results from psychological tests, how many meetings/workshops they attended, which university requirements they satisfied, and much more. The hypothesis is that a combination of certain attributes of the student will be predictive as to whether or not they passed the preliminary year at the University of Massachusetts - Dartmouth.

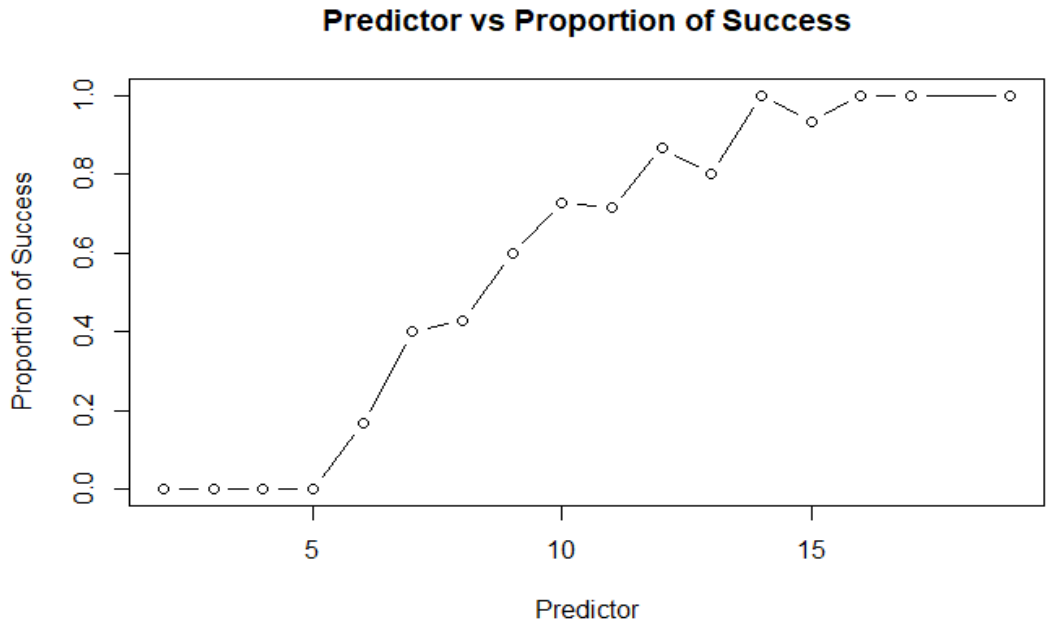
Methods

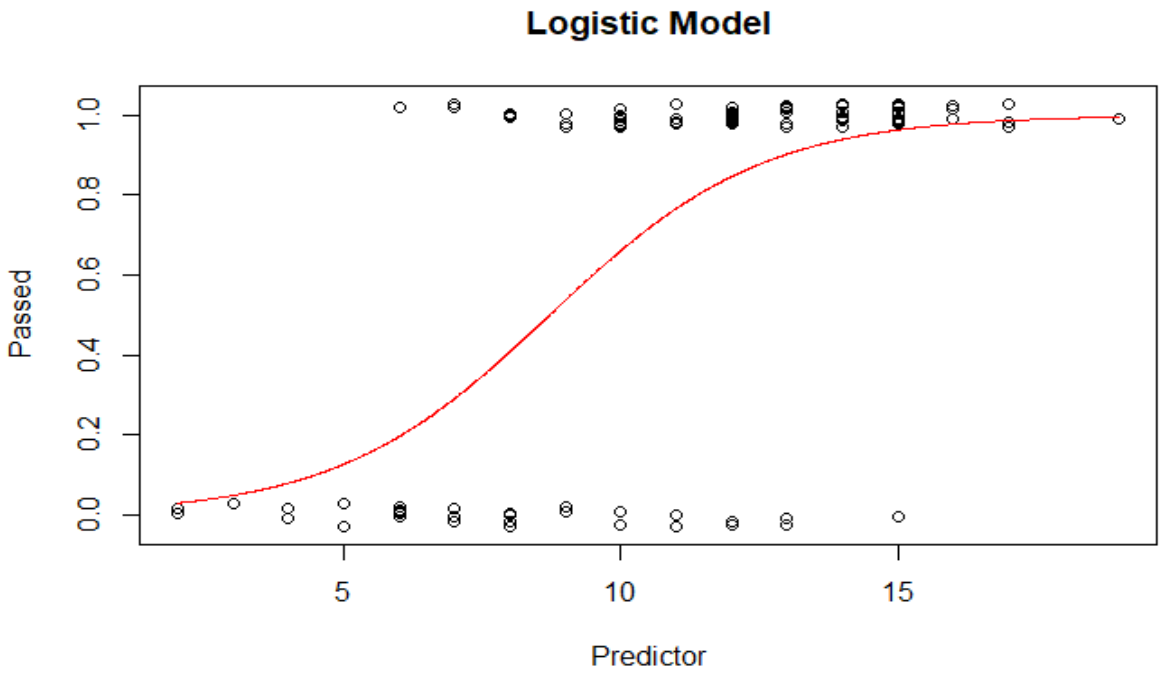
To test the hypothesis that a certain combination of a student's attributes is highly predictive of them passing the preliminary year, some educated guessing will be required. We will first pick some variables that we believe make the most sense in being closely related to whether or not a student passes the preliminary year. Then, a logistic model and multinomial logistic model will be constructed and plotted to measure how accurately these attributes predict a student passing by observing how frequently the model correctly predicts whether a student will pass or fail. Finally, the goodness of fit of the logistic curve will be calculated using Hosmer-Lemeshow tests. This process will be repeated multiple times based on which attributes we want to test. The first attributes we will test are based on how active the student was in campus events/requirements as well as the number of advisor meetings, peer mentor meetings, and workshops attended. Next, the psychological evaluations will be tested to see how well they predict passing the preliminary year. Finally, GPA will be tested as a predictor for passing the preliminary year.

Results

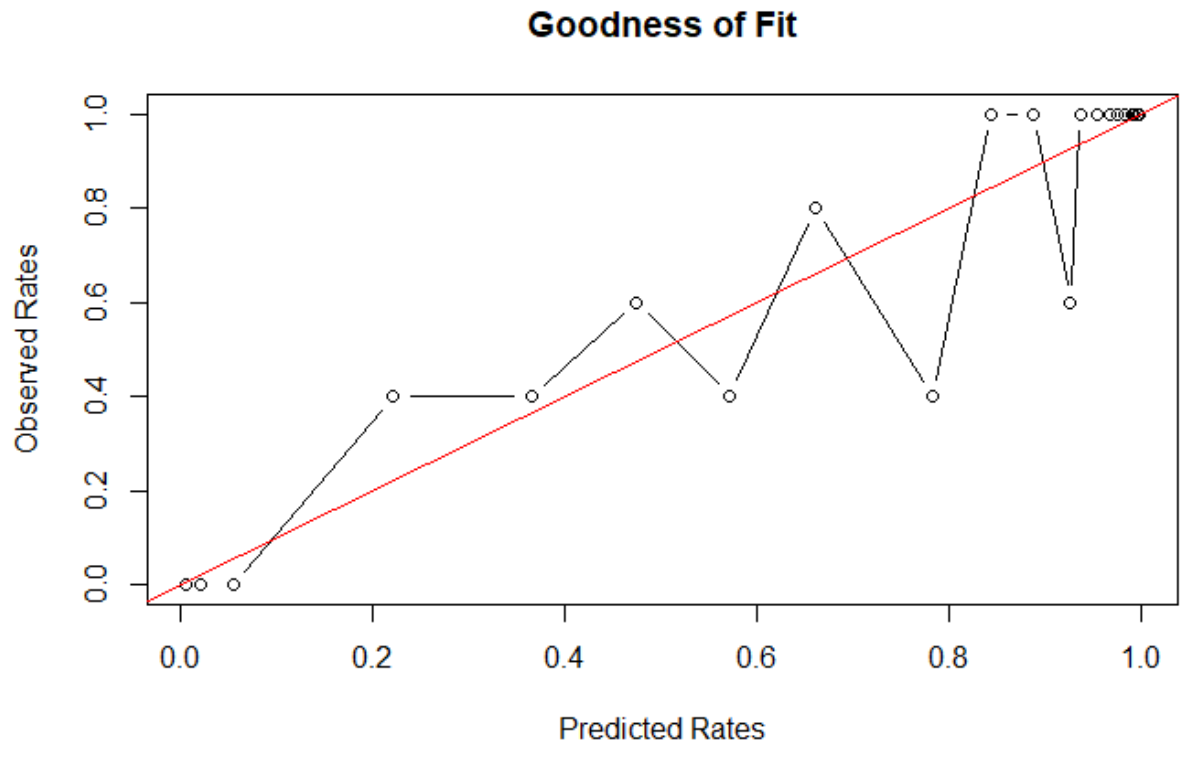
For the first model, we use the sum of the variables: Completed Course, Completed Summer Bridge, Completed Campus Event Requirement, Number of Faculty Advisor Meetings Attended, Number of Peer Mentor Meetings Attended, and Number of Workshops Attended as the Predictor:

Predictor	N	Success	Proportion Success
2	2	0	0
3	1	0	0
4	2	0	0
5	3	0	0
6	6	1	0.1666667
7	5	2	0.4
8	7	3	0.4285714
9	5	3	0.6
10	11	8	0.7272727
11	7	5	0.7142857
12	15	13	0.8666667
13	10	8	0.8
14	9	9	1
15	15	14	0.9333333
16	3	3	1
17	3	3	1
19	1	1	1



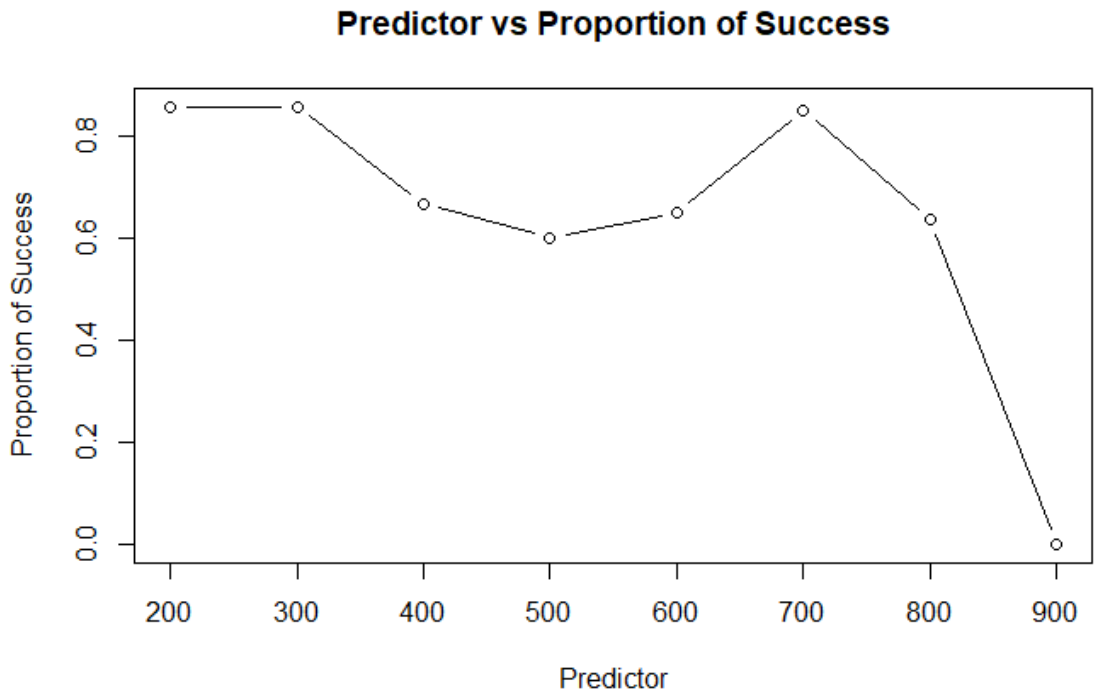


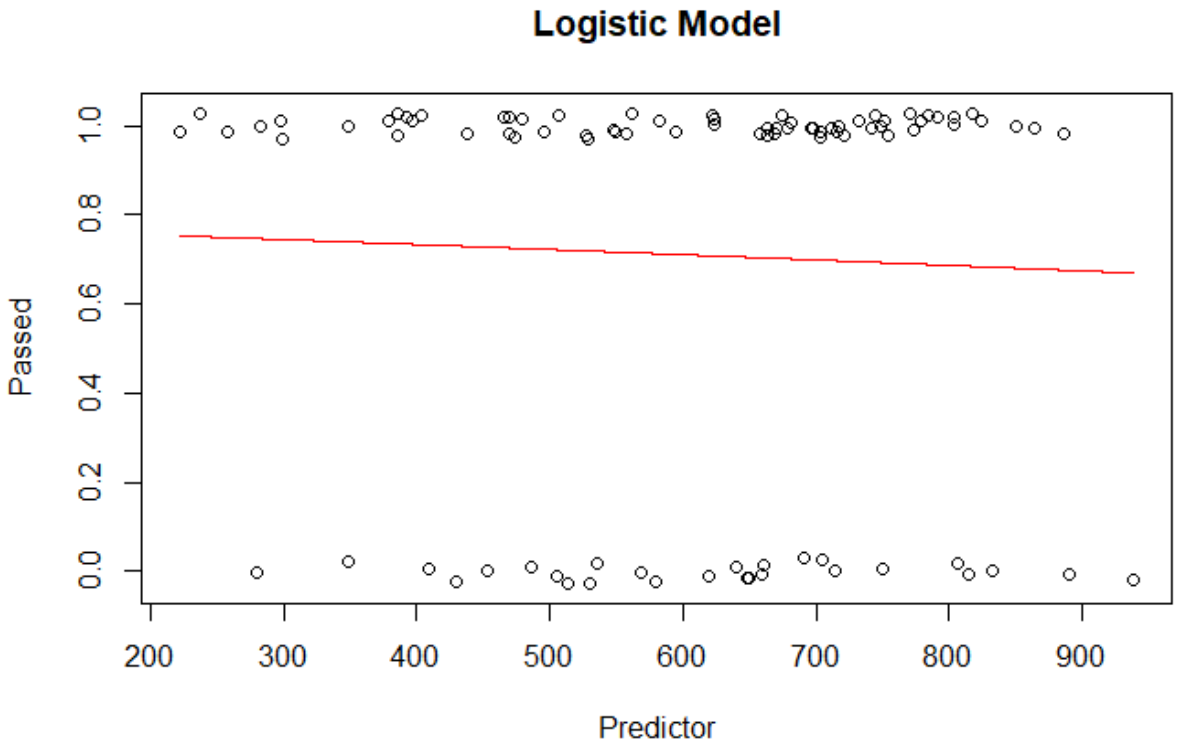
We find that the model predicted approximately 84% of the students completion status correctly and fits the model with a Pearson's R-Squared value of 0.86.



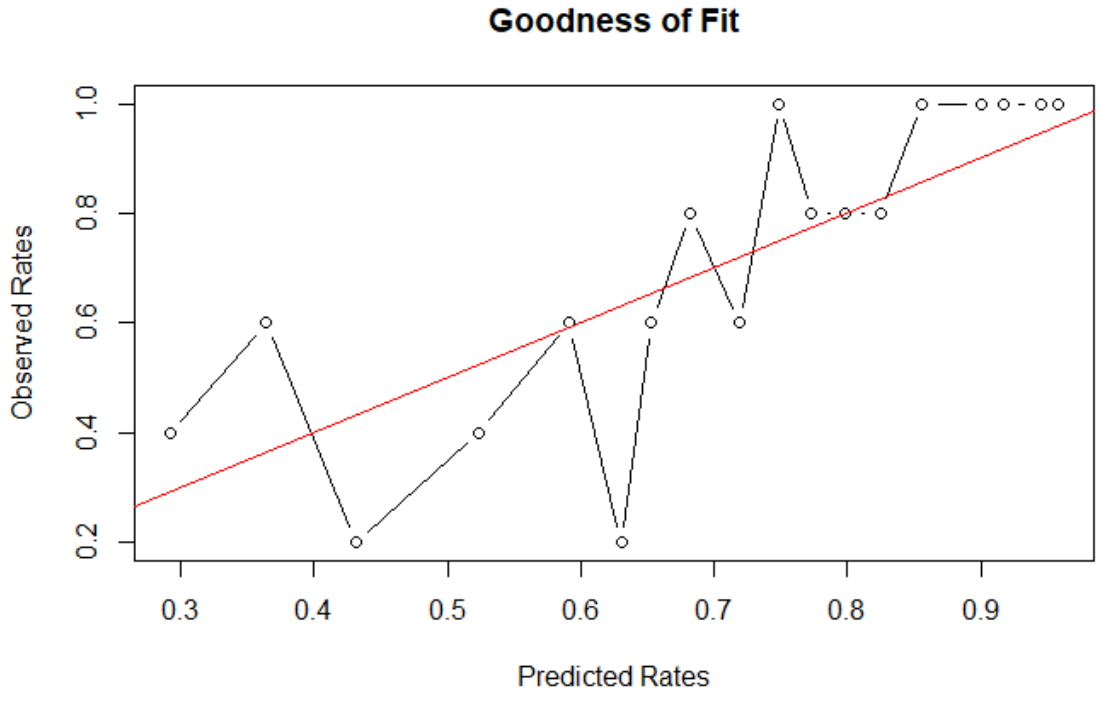
Next, psychological attributes were analyzed: Dropout Proneness, Predicted Academic Difficulty, Educational Stress, Receptivity to Institutional Help, Receptivity to Academic Assistance, Receptivity to Personal Counseling, Receptivity to Social Engagement, Receptivity to Career Guidance, Receptivity to Financial Guidance, Desire to Transfer. The sum of these values will be used as the Predictor.

Predictor	N	Success	Proportion Success
200-299	7	6	0.8571429
300-399	7	6	0.8571429
400-499	12	8	0.6666667
500-599	15	9	0.6
600-699	20	13	0.65
700-799	20	17	0.85
800-899	11	7	0.6363636
900-999	1	0	0



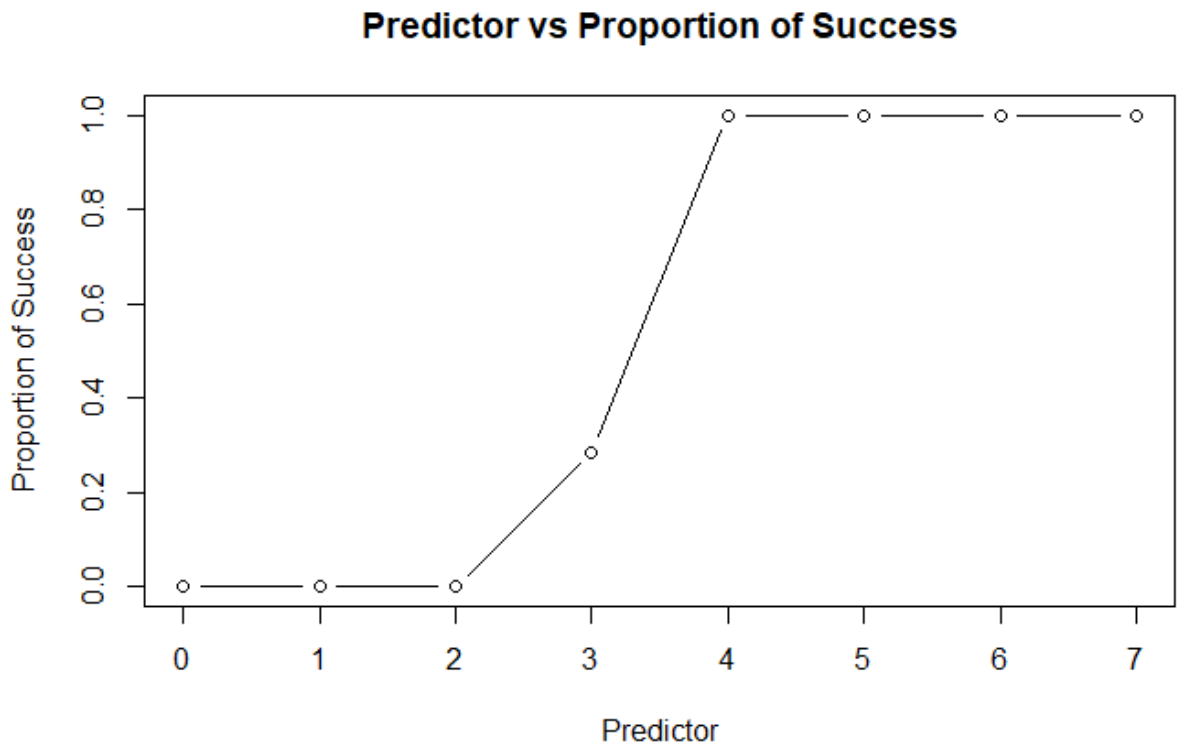


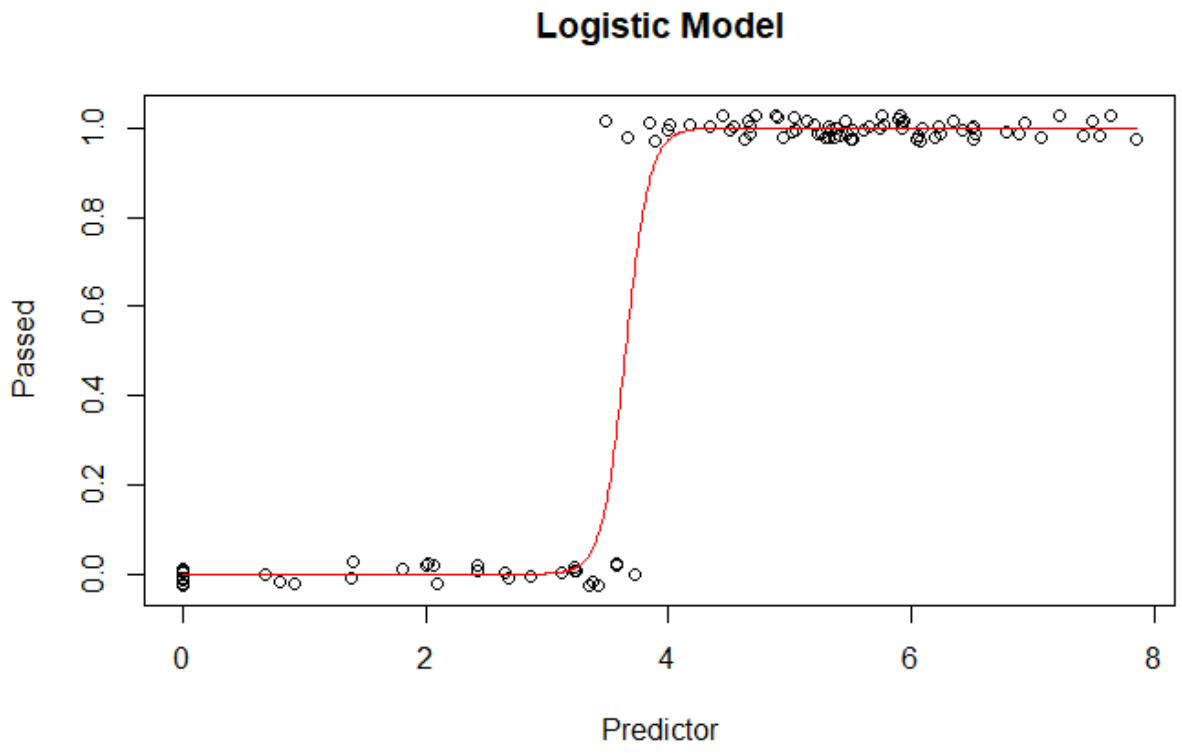
We see that the logistic model is a very poor fit when summing up all the values, yet when using the multinomial logistic model, which considers each value independently, we find the model predicted approximately 74% of the students completion status correctly and fits the model with a Pearson's R-Squared value of 0.66.



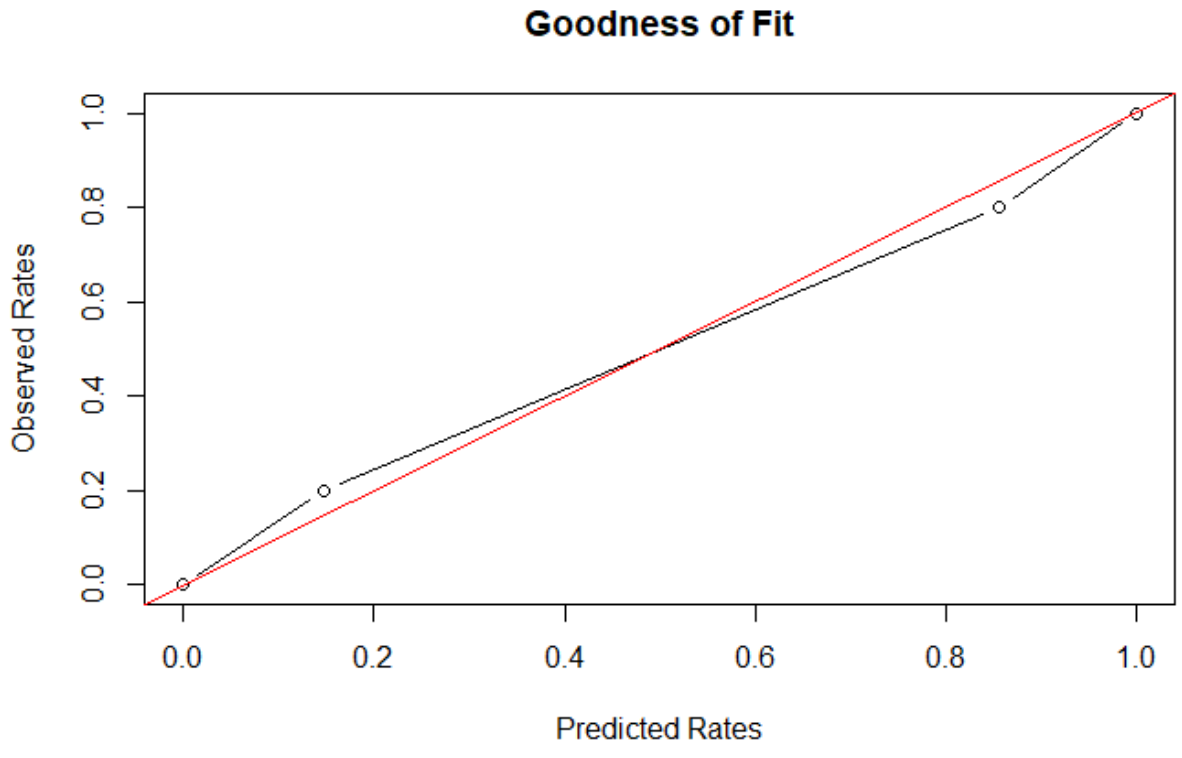
Finally, we will use the sum of students Fall Semester GPA and Spring Semester GPA to predict whether or not they passed.

Predictor	N	Success	Proportion Success
0.00-0.99	12	0	0
1.00-1.99	3	0	0
2.00-2.99	9	0	0
3.00-3.99	14	4	0.2857143
4.00-4.99	16	16	1
5.00-5.99	30	30	1
6.00-6.99	16	16	1
7.00-7.99	7	7	1





The Logistic Model, when considering the sum of the Fall Semester GPA and Spring Semester GPA, only incorrectly predicted the results for 2 students. The Multinomial Logistic Model, which looked at each semester individually, only incorrectly predicted the results for 1 student. The Pearson's R-Squared value for the Goodness of Fit for the Logistic Model is 0.999.



Conclusion

The goal of this experiment was to determine if a particular combination of attributes of a student could predict whether or not they would pass a preliminary year of college. Three different features of the student were observed: their activities during the semester, the results of their psychological exams, and their GPA at the end of the semester. The psychological tests did not predict the results of the student's passing very well, even though that's exactly what they were designed to do. Using a multinomial logistic model, the psychological tests correctly predicted the results of 74% of students. This isn't horrible, but probably not as high as it should be considering what the tests were designed to do. These results don't necessarily mean that psychological tests are a bad method at predicting whether a student will pass or fail, but perhaps in this case the wrong tests are being used. Using the student's activities during the semester, such as how many workshops they visited and how many advisor meetings they attended, predicted 84% of the student's results correctly. As expected, their GPA had the best correlation with passing the preliminary year. Though, this is not exactly a revolutionary discovery, as it's typically the individual's grades that determine whether or not they pass in the first place. The right question to ask would be, what features are common in students with high GPAs? Based on these results, it makes sense that students who attend more workshops, advisor meetings, etc, will increase their odds of passing, but won't guarantee it. Their performance in class is ultimately what decides whether or not they pass, so how they use their time and energy, in addition to their academic motivation will likely be a better predictor of their success.