Will Gillette

Prof. Grossbauer

STAT-242: Statistics 2

9 April 2023

Validity of SAT Score as a Predictor of College Success: Relevancy and Value in College Admissions Decisions

William Gillette[1]

Abstract

In the admissions process, colleges consider several measures in assessing a student's readiness as well as likelihood of success at the college level. For close to one hundred years, colleges in the United States have relied heavily on standardized testing scores, primarily the SAT, as a predictor of college readiness and success. This research topic focused on determining the validity of the SAT as a predictor of college success (as measured by First Year Grade Point Average), how the SAT measured against other admissions criteria, and if and how the SAT's validity has changed over time. To do this, various analyses were conducted, including multiple linear regression analyses, a series of two-sample t-tests, and a Chi-Squared test to analyze a dataset from Educational Testing Service (ETS) consisting of SAT and Grade Point Average (GPA) data from 1996 for 1,000 students at an undisclosed college. Also reviewed were results from several research studies, ranging from the years 2006 to 2017, regarding SAT score as a predictor of college success. Analyses of the ETS data were conducted in R and showed a

[1] William Gillette, third year student at Ursinus College, Statistics 2 class.

moderately strong positive correlation between a student's cumulative SAT score and First Year GPA (FYGPA). However, findings showed an even stronger positive correlation between High School GPA (HSGPA), another predictor of college success, and FYGPA. Finally, a review of data from numerous studies over time showed consistent results that HSGPA was a better predictor of SAT score and that combined use of these measures was the best predictor of college success, as measured by FYGPA. Ultimately, while SAT has been shown to be a strong predictor of college success, studies suggest that HSGPA is a better predictor even when viewed over time and across different sample sizes.

Introduction

Colleges rely heavily on standardized test scores, primarily the SAT, as a part of their decision-making for the admissions process. As a result, students, parents, and educators place great emphasis on these test scores, spending considerable time and sometimes money to help students achieve their greatest scores. Some colleges, like Ursinus College, have been test-optional for an exceptionally long time. However, more recently, many colleges, including Ivy League (e.g., Columbia University), are questioning the validity of SAT scores as predictors of college success as measured by First Year GPA. Given this doubt and an increasing change in college admissions policy, it is important to evaluate whether the SAT is valid as an accurate predictor of college success. Furthermore, it raises the question as to whether there are better predictors that should be considered . The College Board, a nonprofit organization that develops and administers the SAT, has argued that its test is a critical predictor and should continue to be used for college admissions decisions. It is necessary to assess results that either support or dispute this to inform all who are involved with college admissions: students, parents, teachers, and admissions officers.

This research topic sought to address three primary questions. First was to determine the validity of SAT score as a predictor of college success, as measured by FYGPA. This was assessed by reviewing data from 1996 provided by Educational Testing Services (ETS) for 1,000 students. The second primary research question was to determine how the SAT compares to other measures in predicting college success. To determine this, observations from the ETS dataset for High School GPA were analyzed. Lastly, the third primary research question was to determine whether the SAT’s predictability for college success has changed over time - spanning from 1995 to 2017. This was done by reviewing results from numerous studies and comparing them to the ETS data. Based on the available data, one secondary research question was added using the ETS data, segmented by sex, to determine if the relationships found in primary research questions one and two held true when the data was cut this way.

The ETS dataset included 1,000 observations randomly sampled by the College Board from an undisclosed college. Each observation contains six variables: the individual’s High School GPA, First Year GPA, cumulative SAT score, verbal SAT section score, math SAT section score, and sex. This data was collected directly from SAT tests taken in 1996. Data from other studies was similarly collected with high school GPA being reported by the student in their profile when taking the SAT. FYGPA in all datasets was an official GPA provided by colleges.

Methods

Sample: Data used was for 1,000 students (both males and females) who took the 1996 SAT test and later enrolled in an unnamed college.

Data and Variables: Data for the cumulative SAT score and its components came directly from The College Board. ETS obtained data identifying sex and high school GPA directly from the student taking the test from their SAT registration form. Data for FYGPA was obtained directly from this unnamed college.

Procedures:

To assess the strength of the SAT and high school GPA as predictors of college success, which corresponds to research questions one and two, linear regression analyses were conducted. The distribution of data points in the scatterplot between SAT score and First Year GPA had a linear shape, verifying that linear regression analysis is appropriate for this sample. Similarly, the scatterplot of the relationship between High School GPA and First Year GPA was also linear, so a linear regression analysis for those variables was also conducted. The slopes and intercepts of the regression lines for the respective distributions were analyzed. The respective correlation coefficients were also obtained and analyzed as well to give insight into the strength and direction of the correlations between these variables. Finally, analyses were made in the plots between these variables and their residuals to determine whether trends exist between them.

In addition to considering SAT scores in the admissions process, colleges tend to award scholarships based on a defined SAT score, anticipating that a student achieving such a score is likely to achieve a Dean’s List qualifiable GPA during their first year of college, measured as a GPA of 3.5 or above. Ursinus College, for example, awards a $35,000 scholarship to anyone who achieves a cumulative SAT score of 1250 or higher. Therefore, results regarding this relationship are valuable in informing this process and policy and tie to question one concerning the validity of SAT score predicting college success. To assess the claim that students who score

1250 or higher on the SAT are highly likely to achieve a Dean's List GPA, the ETS data was partitioned into two explanatory groups and two response groups for a Chi-Squared analysis. The explanatory groups consisted of individuals who achieved below a 1250 SAT score and those who achieved a 1250 SAT score and above. The response groups consisted of individuals who achieved a Dean's List GPA (3.5 or higher FYGPA) and those who did not. The aim of this test was to determine whether the proportion of individuals who scored below a 1250 SAT score and achieved a Dean's List GPA differed from that of the group with individuals who scored 1250 or higher. This aids in assessing the strength of the SAT as a predictor of college success to answer the first research question.

To address research question 3, a review of several published studies was conducted. The purpose was to determine if SAT score results and the relationship between cumulative SAT score and first year GPA was consistent over time since the format and content of the SAT exam has changed over this period (1995-2017). To accomplish this, comparisons were made between the correlation coefficients from the linear regression models developed from the ETS data and data from several outside studies. These studies used similar procedures to the ETA analysis to determine the correlations between cumulative SAT score and FYGPA and High School GPA and FYGPA. These additional studies have dates and sample sizes that differ from the ETS data and are thus powerful resources from which to make comparisons to determine whether the SAT's predictability for college success has changed over time, and to see if and how sample size affected results.

Finally, to tackle the secondary research question, the ETS data was partitioned by sex to determine if males and females differ in any of the primary variables: FYGPA, HSGPA, and cumulative SAT score. Three two sample t-tests were conducted between sex (Male, Female) and the primary variables (HSGPA, cumulative SAT score, and FYGPA). This is important in determining whether the same trends occur between the sexes with all these variables or if one metric (i.e., HS GPA or cumulative SAT score) is a better predictor of college success than the other for a certain sex.

Results

Table 1 depicts the summary statistics for the ETS dataset. Its variables are all quantitative except for sex, which is a categorical variable with a value of "1" indicating a male student and that of "2" indicating a female student. The variables sat_v and sat_m consist of values between 200 and 800 and represent the verbal and math section scores, respectively. Additionally, the variables hs_gpa and fy_gpa consist of values between 0 and 4.0 and represent the high school GPA and first year GPA, respectively. Finally, the sat_sum variable represents the cumulative SAT score and ranges from 400 to 1600.

Table 1: Summary Statistics of ETS Data

column <chr>	n <dbl>	mean <dbl>	sd <dbl>
sex	1000	1.48400	0.4999940
sat_v	1000	489.34000	82.3392017
sat_m	1000	543.95000	84.5011121
sat_sum	1000	1033.29000	142.8736812
hs_gpa	1000	3.19810	0.5416475
fy_gpa	1000	2.46795	0.7408052

To address research questions 1 and 2, this study began by analyzing the ETS data to determine the relationship between cumulative SAT score and First Year GPA. In this analysis, sat_sum was compared to fy_gpa by performing a simple linear regression analysis using the entire dataset of 1,000 observations. Thus, the test hypotheses for this analysis were "$H_0: \beta = 0$" and "$H_a: \beta \neq 0$." Figure 1.1 depicts a positive slope for the regression line in the scatterplot of sat_sum against fy_gpa. The equation for the least squares regression line for this model was "$FYGPA = 0.0019 + 0.0024(SAT)$," with a 95% confidence interval of $[0.002, 0.003]$ for the slope. The observed t-statistic for the slope was $t = 16.39$ with a corresponding p-value of $p < 2 * 10^{-16}$. The observed correlation coefficient for the cumulative SAT score and first year GPA was $r = 0.46$, and the observed R-squared value was $r^2 = 0.2119$. A scatterplot of the cumulative SAT score against the respective residuals is displayed in Figure 1.2.

Figure 1.1: Scatter Plot of Cumulative SAT Score against First Year GPA

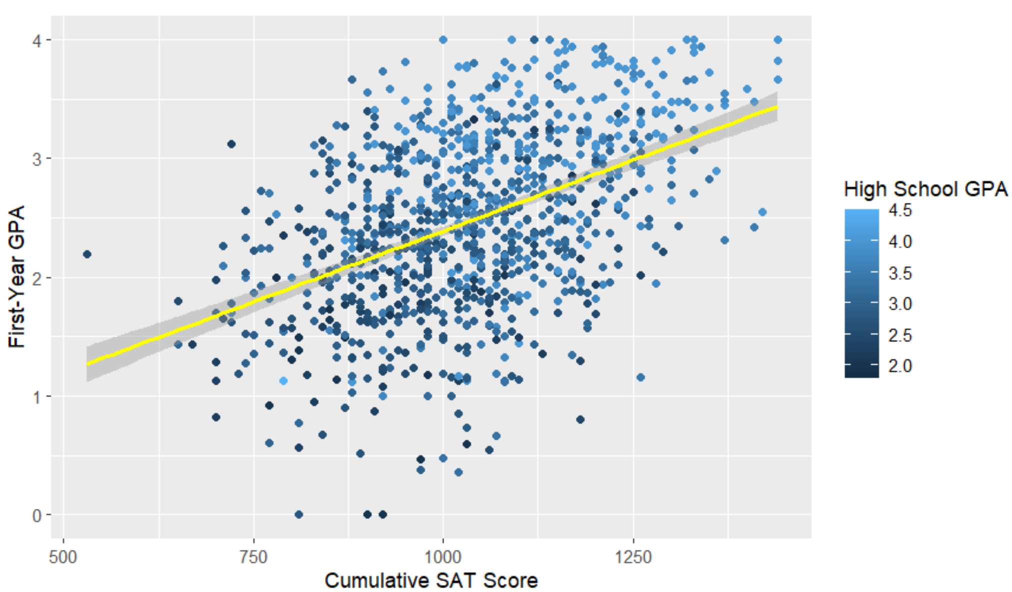
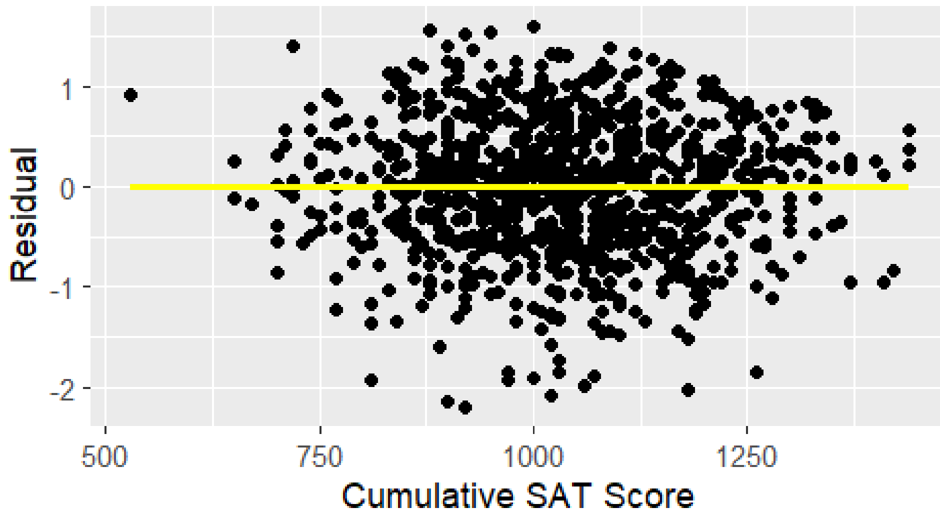


Figure 1.2: Residual Plot of Cumulative SAT Score



Conducting a similar simple linear regression analysis for high school GPA with the same test hypotheses, it was found that the regression line had a positive slope, as shown in the corresponding scatter plot in Figure 2.1. The equation for the least squares regression line for this model was "$\widehat{FYGPA} = 0.09132 + 0.74314(HSGPA)$." The observed t-statistic for the slope was $t = 20.447$, with corresponding p-value of "$p < 2 * 10^{-16}$." The respective confidence interval for the observed slope was $[0.672, 0.814]$. The observed correlation coefficient was $r = 0.54$, and the R-squared value was $r^2 = 0.2952$. A scatterplot of the High School GPA against the respective residuals is displayed in Figure 2.2.

Figure 2.1: Scatter Plot of High School GPA against First Year GPA

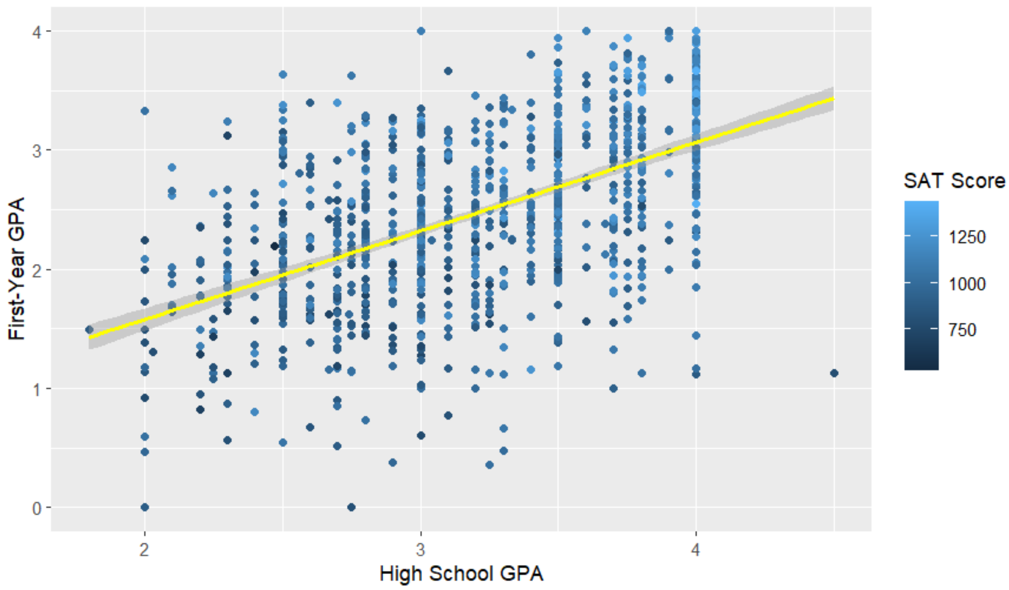
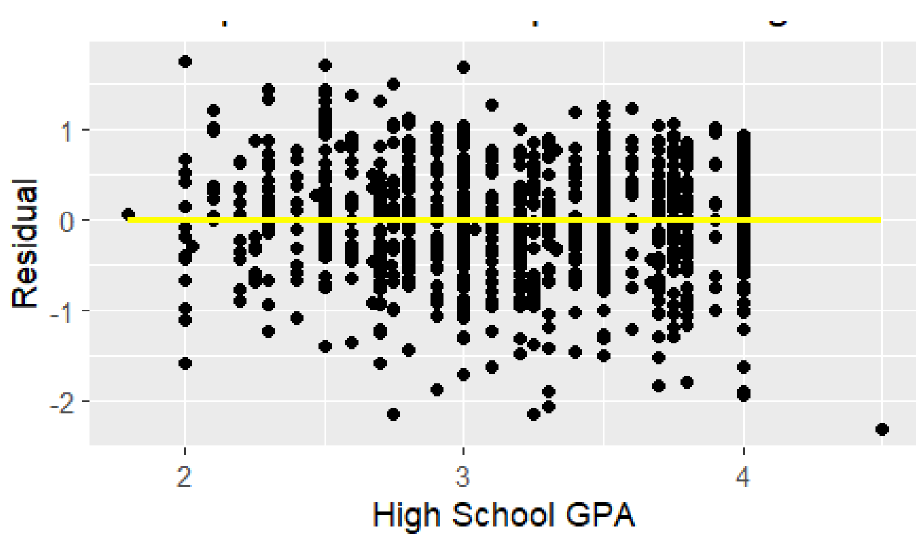


Figure 2.2: Residual Plot of High School GPA



A multiple regression analysis with independent variables SAT and HSGPA and dependent variable FYGPA was also conducted with the hypotheses, "H_0: $\beta_0 = \beta_1 = 0$" and "H_a: At least one of the coefficients does not equal zero", where β_0 and β_1 represent the population HSGPA and cumulative SAT score coefficients respectively. The regression line equation for this model was: "$\widehat{FYGPA} = -0.873 + 0.0014(SAT) + 0.579(HSGPA)$." In this analysis, the coefficients for sat_sum and hs_gpa had respective t-statistics of 9.878 and 15.069, sharing a p-value of "$p < 2 * 10^{-16}$." The confidence intervals for these coefficients were $[0.001, 0.002]$ and $[0.504, 0.655]$ respectively. The observed adjusted R-squared value was $r^2 = 0.357$ with a corresponding F-statistic of $F = 278.1$ on two and 997 degrees of freedom, having a p-value of "$p < 2.2 * 10^{-16}$." Table 2 shows the correlation matrix for the variables in this analysis, displaying correlation coefficients of 0.46 for SAT score and FYGPA, 0.54 for HSGPA and FYGPA, and 0.43 for SAT score and HSGPA. Figure 3.1 shows the three-dimensional scatter plot of HSGPA and SAT score against FYGPA, while Figures 3.2 and 3.3 depict the spread and distribution of the standardized residuals from this analysis. Finally, Figure 3.4 shows the Cook's Distance Plot for this analysis.

Table 2: Correlation Matrix for the Variables in the MLR Analysis

	fy_gpa	sat_sum	hs_gpa
fy_gpa	1.00	0.46	0.54
sat_sum	0.46	1.00	0.43
hs_gpa	0.54	0.43	1.00

Figure 3.1: Scatter Plot of HSGPA and SAT Score against FYPGA

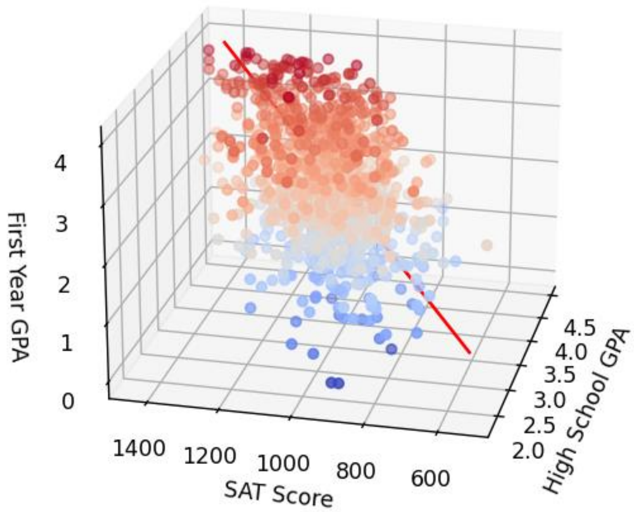


Figure 3.2: Standardized Residual Plot for Predicted FYGPA in the MLR Analysis

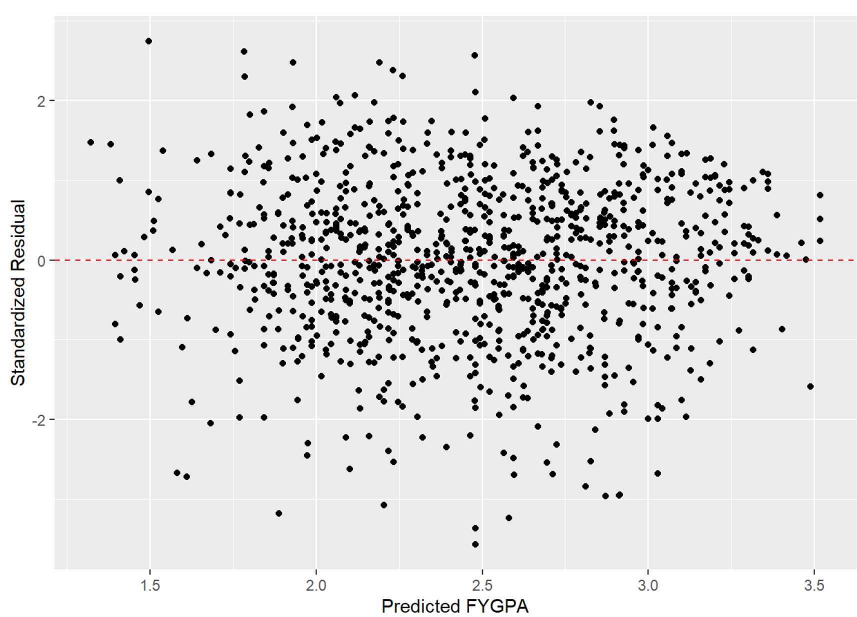


Figure 3.3: Standardized Residual Histogram for Predicted FYGPA in the MLR Analysis

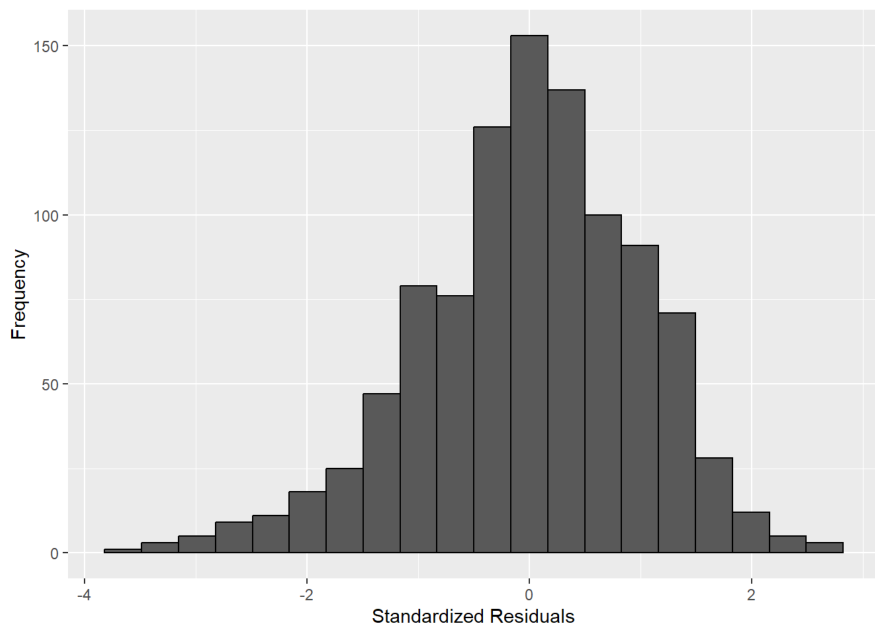
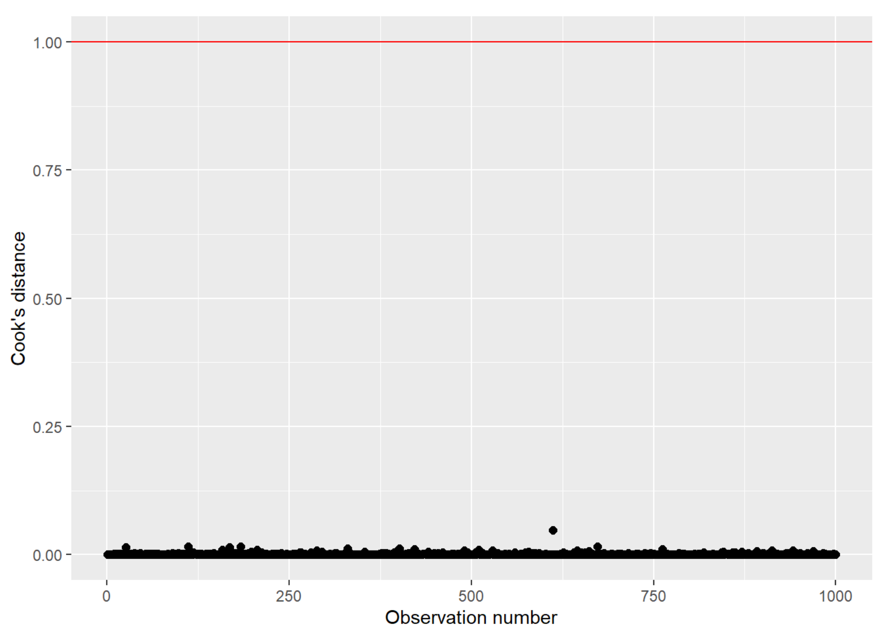


Figure 3.4: Cook's Distance Plot for the MLR Analysis



Lastly, to round out the analyses for the first two research questions, the results of the Chi-Squared test were used to examine proportions based on SAT score and Dean's List GPA are depicted in Figures 4.1, 4.2, and 4.3. These bar charts depict certain proportions based on these groups in the sample. The observed proportions of students who achieved a 3.5+ FYGPA were $\hat{p}_{1250+} = 0.299$ and $\hat{p}_{<1250} = 0.0654$ for the 1250+ SAT Score and Below 1250 SAT Score groups, respectively.

Figure 4.1: Proportion of Students who Achieve a Dean's List First Year GPA

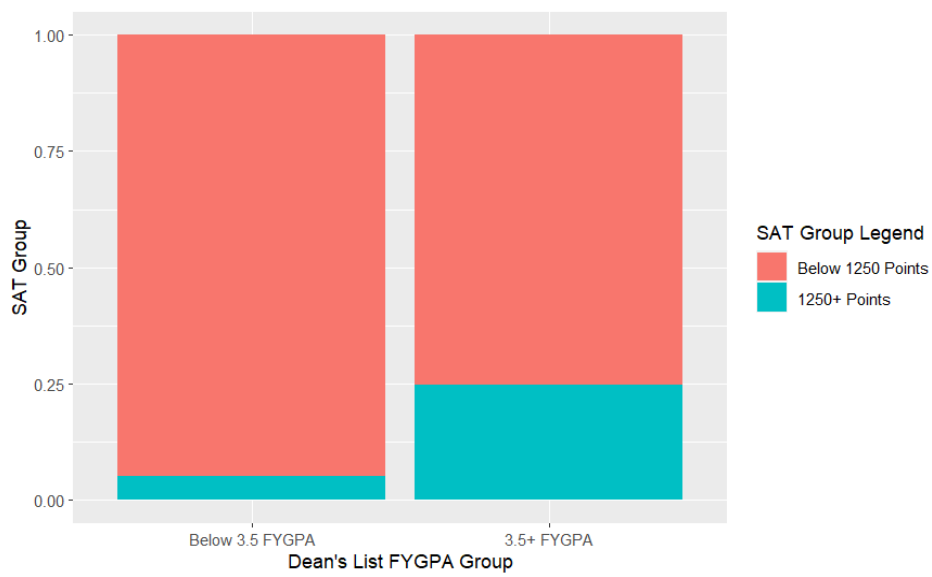


Figure 4.2: Proportion of Groups who Achieve a Dean's List First Year GPA

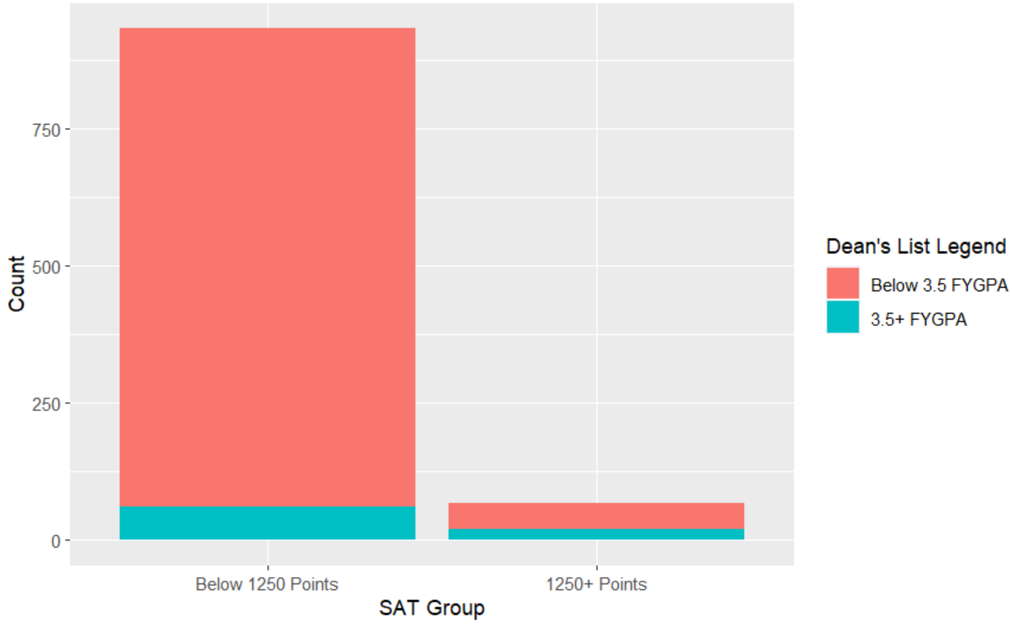
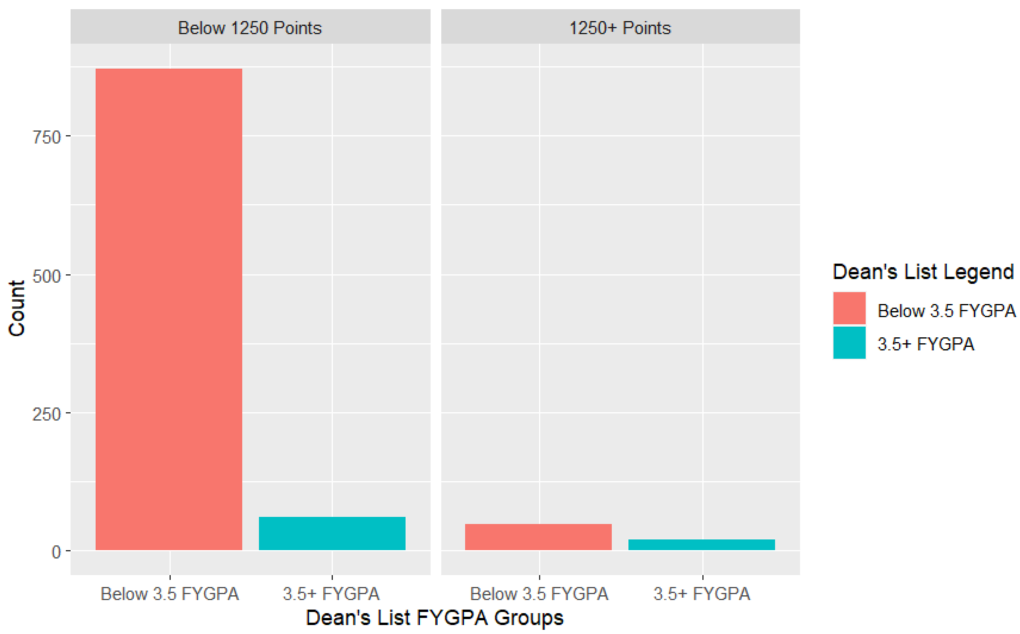


Figure 4.3: Faceted Bar Charts of Dean's List First Year GPA Counts in SAT Groups



As shown in Table 3.1, the validity conditions for a Chi-Squared test of significance are met because there are at least ten observations in each group. The observed Chi-Squared value was $X^2 = 42.561$ with one degree of freedom and a corresponding p-value of $p = 6.85 * 10^{-11}$.

Table 3.2 depicts the expected counts in each group assuming that there is no association between achieving a 1250+ SAT score and attaining a Dean's List FYGPA.

Table 3.1: 1250 SAT Score vs. Dean's List GPA Two-Way Table

	3.5+ FYGPA	Below 3.5 FYGPA
1250+ Points	20	47
Below 1250 Points	61	872

Table 3.2: Expected Counts of 1250 SAT Score vs. Dean's List GPA Two-Way Table

	3.5+ FYGPA	Below 3.5 FYGPA
1250+ Points	5.427	61.573
Below 1250 Points	75.573	857.427

Analysis was then conducted to address research question 3 which centered around the validity of SAT score as a predictor over time.; This was relevant since the test has undergone several changes over time and results were available through a review of various published studies. Results of the assorted studies are summarized in Table 4.

Table 4: Correlation Coefficients for SAT, HSGPA vs. FYGPA for Various Studies

Study	SAT to FYGPA	SGPA to FYGPA	Combined	Sample Size (N)	Year
Patterson, et al.	0.54	0.55	0.62	150,377	2006
Westrick, et al.	0.51	0.53	0.61	223,858	2017
ETS	0.46	0.54	N/A	1,000	1996

In the study conducted by Patterson, et al., the data sample consisted of approximately 150,000 students who took the SAT in 2006 and results were HSGPA having a correlation coefficient of $r = 0.55$ and SAT of $r = 0.54$ (37). Similarly, in the study by Westrick, et al. (11), which used data from 2017 for 223,000 students across 171 colleges and universities, including Ursinus, results showed a higher correlation coefficient between HSGPA and FYGPA ($r = 0.53$) than between SAT scores and FYGPA ($r = 0.51$). Lastly, as noted previously, the ETS study reported comparable results with an SAT to FYGPA correlation coefficient of 0.46 and a HSGPA to FYGPA correlation coefficient of 0.54.

Results were also available for studies that combined HSGPA with SAT score (Table 4). The study by Patterson, et al. reported a combined correlation coefficient of 0.62 (41) and the study by Westrick, et al. one of 0.61 (11). The combined correlation coefficient is not available for ETS.

Lastly, results in Table 4 show that, over time in the years ranging from 1996 to 2017, the correlation coefficient for SAT and FYGPA remained consistent, ranging from 0.46 to 0.54. Similarly, the correlation coefficient between HSGPA and FYGPA remained consistent, ranging from 0.53 to 0.55.

Given the available data in this study, the opportunity for a secondary research question which examined data by sex arose. The results of the secondary research question which looked at the ETS data for the primary variables (FYGPA, HSGPA, and cumulative SAT score) by sex was that, for each of these comparisons, the histograms between the sexes had remarkably similar shapes. As shown in Figure 5.1, when comparing First Year GPA between the sexes, there was a slight left skew in both groups. The corresponding box plots in Figure 5.2 align with this observation as seen by the dots present beyond the lower tail. Furthermore, the median, lower quartile, and upper quartile for first year GPA in females was slightly higher than that in the males. The data also showed that the mean first year GPA in the female group was higher than that of the male group in this sample ($\overline{x_{female}} = 2.545$ vs. $\overline{x_{male}} = 2.396$).

Figure 5.1: Faceted Histogram of First Year GPA vs. Sex

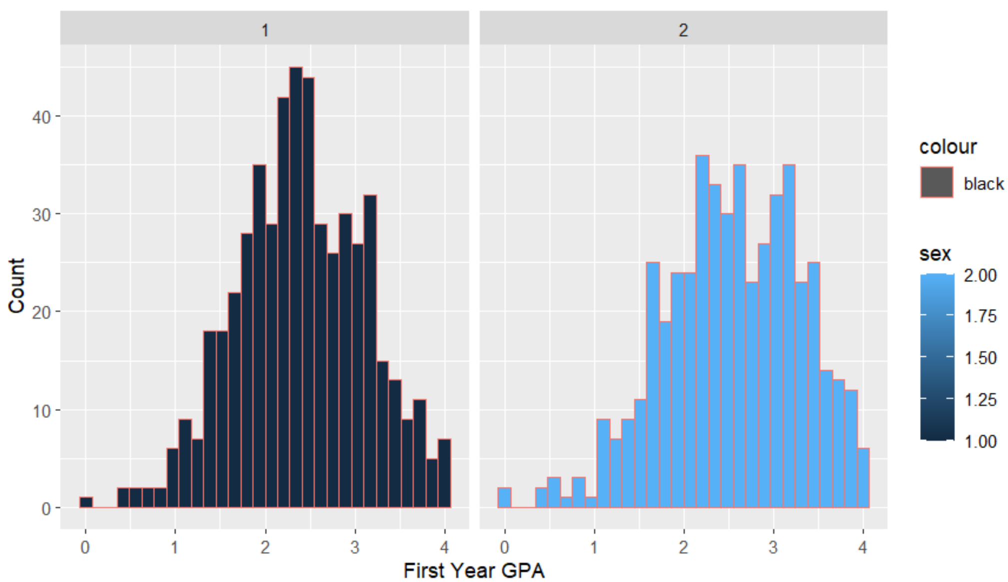
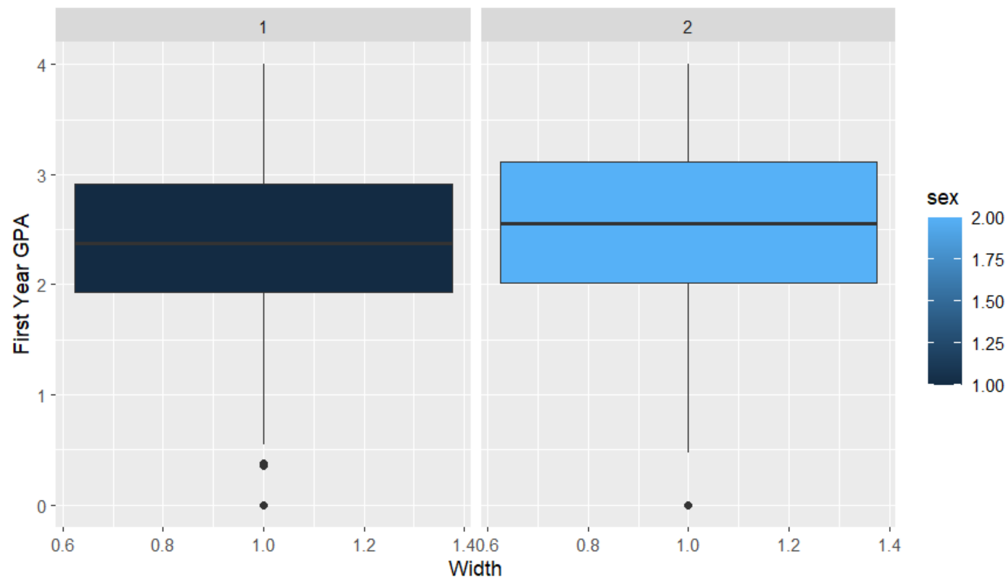


Figure 5.2: First Year GPA Box Plots of Each Sex



As seen in Figure 6.1, the faceted histograms for high school GPA and the sexes have extremely similar shapes. The corresponding box plots in Figure 6.2 depict the same trend in Figure 5.2: the median and upper quartile high school GPAs in the female group were higher than those of the male group. The mean HSGPAs of the female and male groups were $\overline{x_{male}} = 3.277$ and $\overline{x_{female}} = 3.124$ respectively. Additionally, there were no outliers in the distribution of high school GPAs for either of the sexes, but there was a slight left skew in the female group.

Figure 6.1: Faceted Histogram of High School GPA vs. Sex

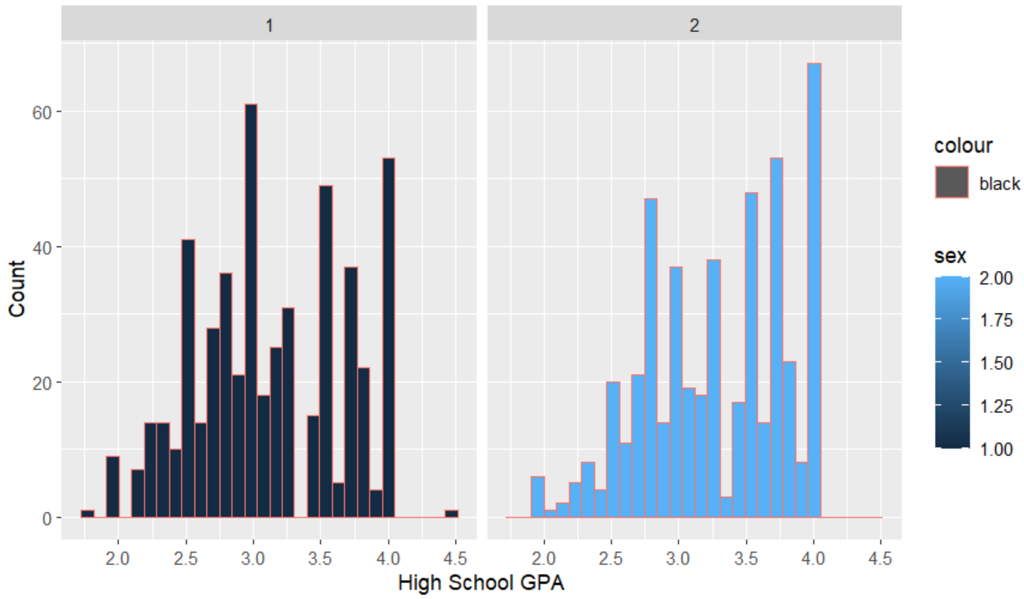
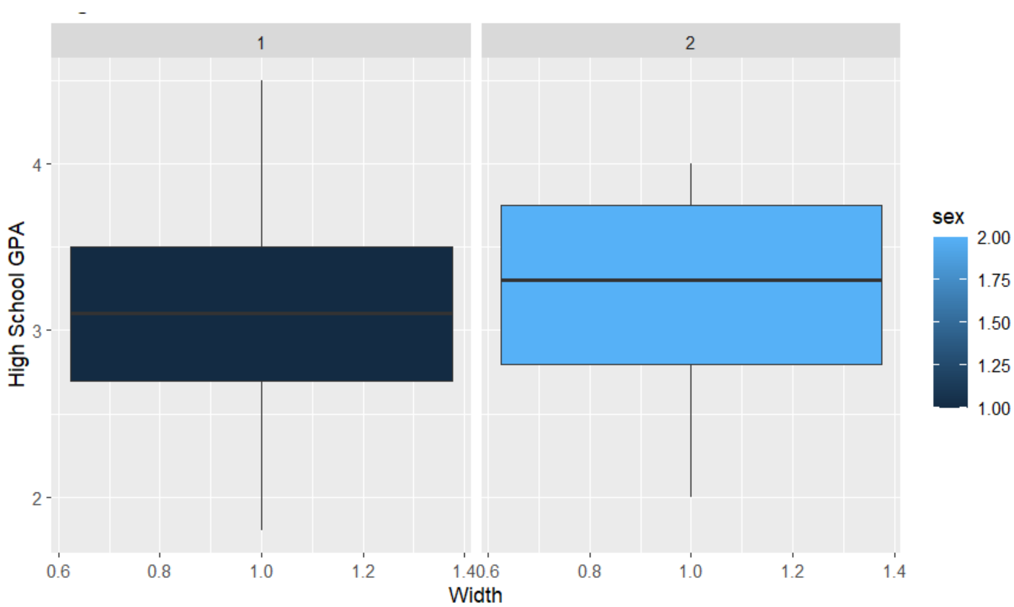


Figure 6.2: High School GPA Box Plots of Each Sex



Furthermore, each sex's cumulative SAT score distribution, as shown in Figure 7.1, appeared to be normally distributed but had a couple of outliers in both directions. Interestingly, as seen in the corresponding box plot in Figure 7.2, the median, lower quartile, and upper quartile SAT scores in the male group were higher than that of the female group.

Figure 7.1: Faceted Histogram of Cumulative SAT Score vs. Sex

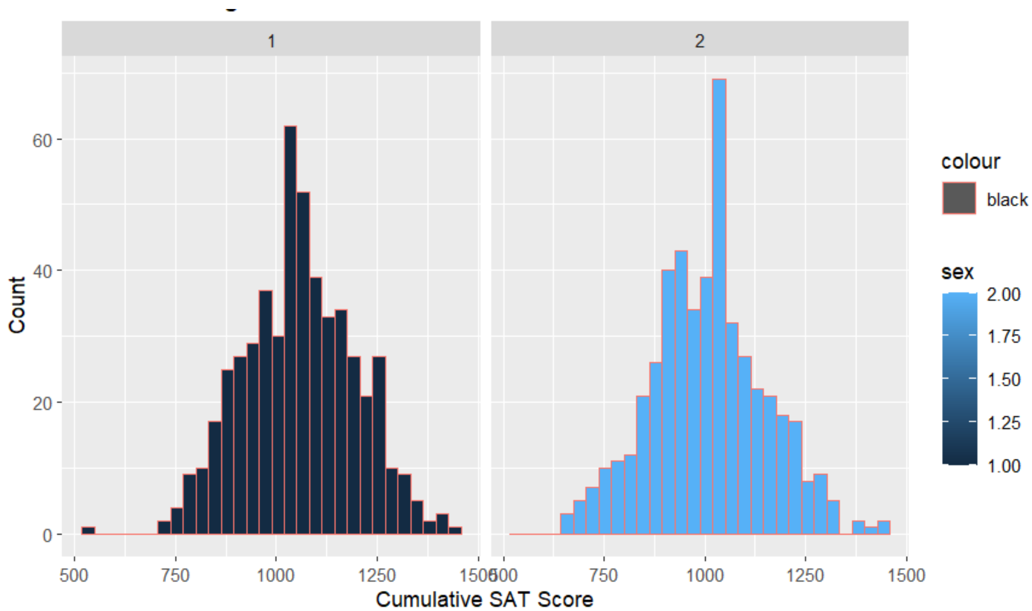
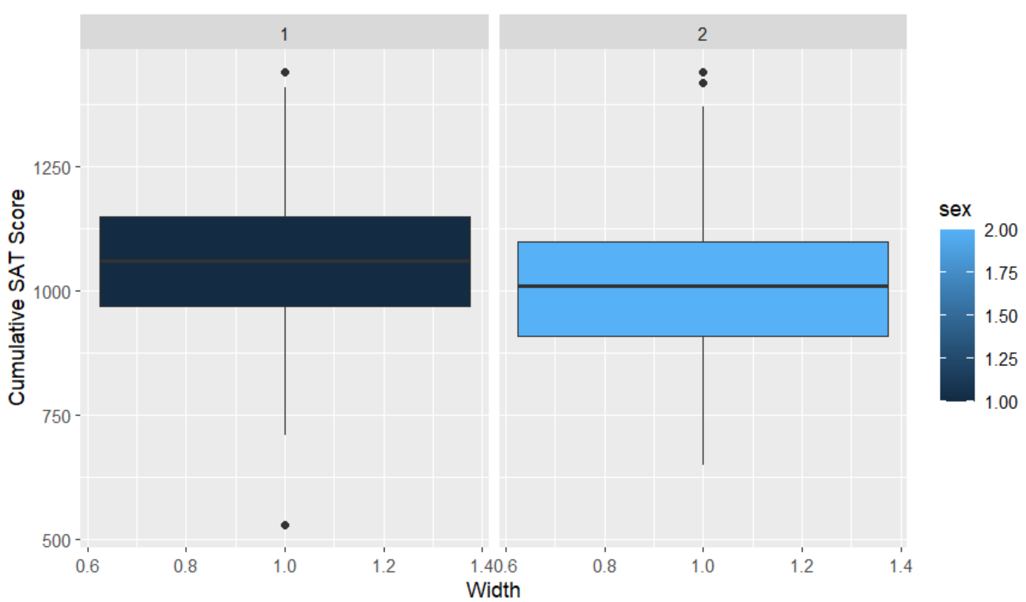


Figure 7.2: Cumulative SAT Score Box Plots of Each Sex



The two-sample t-test for first year GPA resulted in an observed t-statistic of $t = 34.81$ with a corresponding p-value of $p = 2.2 * 10^{-16}$ and 95% confidence interval of $[0.9285, 1.039]$. A similar test for cumulative SAT score resulted in an observed t-statistic of $t = 228.37$ with a corresponding p-value of $p = 2.2 * 10^{-16}$ and a 95% confidence interval of $[1022.94, 1040.67]$. A final test on high school GPA resulted in an t-statistic of $t = 73.53$ with a corresponding p-value of $p = 2.2 * 10^{-16}$ and a 95% confidence interval of $[1.668, 1.76]$. For all the above tests, the observed t-statistics were all statistically significant.

Discussion

The results of the first test, the simple linear regression analysis conducted on cumulative SAT score and First Year GPA (Figure 1.1), produced a regression line with a positive slope. The observed slope of 0.002 indicates that the model predicts that a one-point increase in the cumulative SAT score results in a 0.002-point increase in the first year GPA. Additionally, because the 95% confidence interval of [0.001, 0.002] for the observed slope does not include the null hypothesis value of zero, the observed slope is statistically significant. Based on the interval, we are 95% confident that as the cumulative SAT score increases by one point, the first year GPA increases by between 0.002 and 0.003 points. The correlation coefficient indicates a moderately strong and positive correlation between SAT score and first year GPA. Furthermore, the R-squared value of 0.212 indicates that 21.2% of the variability in the FYGPA is explained by the linear relationship with the cumulative SAT score, while the remaining 79.8% variability is due to confounding variables. In “All About Accountability / Branded By a Test,” James Popham presents the importance of this statistic. He speaks of the importance of the R-squared value, “...you must *square* that predictive validity coefficient of .50 to get an accurate idea of how predictive an SAT or ACT score really is." and notes that, with the data he used, "This means we can accurately predict only 25 percent of a college student's grades on the basis of that student's SAT or ACT scores. Fully 75 percent of a college student's grade performance is explained by other factors, such as student's motivation or study habits."

The corresponding residual plot in Figure 1.2 showed a symmetrical random scatter about the horizontal regression line with no patterns, satisfying the homoscedasticity and normality requirements for simple linear regression. Also, the scatter plot of SAT score against FYGPA is linear, and each pairing is independent, satisfying the linearity and independence requirements. Thus, we meet the validity conditions for simple linear regression with this analysis.

Because the SAT score values for this study and the other studies reviewed in this analysis were sourced directly from The College Board's records, the integrity of the data is valid. For this reason, there are no obvious needs for improvement in this part of the study.

The second simple linear regression analysis between high school GPA and first year GPA yielded a regression line that had a positive slope of 0.74314 (Figure 2.1). This indicates that the model predicts that as high school GPA changes by one point, first year GPA changes in the same direction by 0.74314. The 95% confidence interval for the slope of the regression line was [0.672, 0.814], indicating that the observed statistic was significantly significant. Furthermore, we are 95% confident that as the high school GPA increases by one point, the first year GPA increases by between 0.672 and 0.814 points. The correlation coefficient indicates a moderately strong and positive correlation between High School GPA and First Year GPA. The R-squared value for this analysis implies that 29.5% of the variability in first year GPAs is explained by the linear relationship with High School GPA, while the remaining 70.5% of variability is due to confounding variables. This is higher than the value in the SAT score simple linear regression analysis, suggesting that, based on the model created from this sample, High School GPA has a stronger correlation with First Year GPA than cumulative SAT score does.

The corresponding residual plot in Figure 2.2 showed a random and symmetric spread of the errors, satisfying the homoscedasticity and normality requirements. Additionally, Figure 2.1 is linear with all its pairings independent of each other, satisfying the linearity and independence requirements. Thus, we meet the validity conditions to conduct this secondary simple linear regression analysis as well.

In reviewing the source of the HSGPA data, there is some question as to the accuracy and the validity of the data. In this study and in other studies, the data was self-reported by the students in their SAT test profile. A way to improve this study and increase confidence in the data would be to source the data from official high school transcripts. However, this may not be feasible or appropriate as this would require significant effort for what may be little or no increase in the accuracy of the data. There is no valid reason a student would desire to report inaccurate data so there is no compelling reason to source official data.

The multiple regression analysis tested if either or both independent variables, cumulative SAT score and High School GPA, had a significant relationship with the dependent variable, First Year GPA. For the ETS dataset, $b_0 = 0.0014$, indicating that when holding High School GPA fixed, the predicted first year GPA increases by 0.0014 points for each additional point on the cumulative SAT score. The 95% confidence interval for the cumulative SAT score coefficient indicated that we are 95% confident that when HSGPA is fixed and the cumulative SAT score increases by one point, the FYGPA increases by between 0.001 and 0.002 points. Similarly, $b_1 = 0.579$, indicating that when holding cumulative SAT score fixed, the predicted first year GPA increases by 0.579 points for each increment of 1.0 in the High School GPA. The

95% confidence interval for HSGPA indicated that we are 95% confident that when HSGPA increases by 1.0, the FYGPA increases by between 0.504 and 0.655 points. The t-statistics of 9.878 for the cumulative SAT score slope and 15.069 for the High School GPA slope are statistically significant, each sharing the same corresponding p-value of $p < 2 * 10^{-16}$. Thus, there is sufficient evidence to suggest that at least one of the coefficients in the multiple regression model does not equal zero, meaning that at least one of the independent variables (HSGPA and cumulative SAT score) has a significant linear relationship with the dependent variable, first year GPA. There was an observed adjusted R-squared value of 0.357, indicating that 35.7% of the variability in First-Year GPA is explained by the linear relationship with both High School GPA and cumulative SAT score. The F-statistic corresponding to this observed R-squared value was 278.1 on two and 997 degrees of freedom, with a respective p-value of "$p < 2.2 * 10^{-16}$." The adjusted R-squared value in this model is superior to both simple linear regression models, suggesting that the best predictor of college success (as measured by first year GPA) would be to consider both High School GPA and SAT score to inform college admissions and related decisions.

The validity conditions to conduct Multiple Linear Regression analyses state that the scatter plots of the independent variables against the dependent variable must have linear shapes, there must not be strong multicollinearity, and the residuals must be normally distributed and homoscedastic. Figures 1.1 and 2.1 show linear relationships between SAT score and FYGPA and HSGPA and FYGPA, satisfying the linearity requirement. As seen in the correlation matrix (Table 2), SAT score and FYGPA have an observed correlation of 0.43 in the ETS dataset, suggesting that their interaction is not strong, meeting the weak multicollinearity requirement. Figures 3.2 and 3.3 depict the plot and histogram for the standardized residuals, showing a random spread and approximately normal distribution, meeting the final two requirements. Furthermore, the Cook's Distance Plot in Figure 3.4 shows that there were no outlier residuals in this sample. Thus, the results from the MLR analysis are accurate as the assumptions are met.

Like the discussion above regarding the validity and accuracy of SAT and HSGPA data, the study may be improved if official HSGPA data was available. However, a review of combined data as is done in this last scenario mitigates the risk that the data is inaccurate, affecting study results.

For the Chi-Squared test, conducted after partitioning the data into four groups based on whether individuals achieved a 1250 SAT Score and 3.5 FYGPA, there were at least ten independent observations in each group, meeting the validity conditions to conduct a Chi-Squared test of significance. The results of the Chi-Squared test indicated that the observed counts in each group of this sample differed significantly from the counts expected if there were no difference in the proportions. Thus, we have sufficient evidence to suggest that there is a difference between the proportions of individuals who achieved a 1250 and above SAT score and a Dean's List GPA and that of the group who achieved below a 1250 SAT score. This indicates that there is an association between whether students achieve a 1250 SAT score and whether they achieve at least a 3.5 GPA during their first year of college.

Since the Chi-squared analysis used official SAT score data, there are no improvements needed for this analysis. However, it may be worthwhile to review other studies that have partitioned data by SAT score to examine results and determine consistency with this study's findings.

In the review of various published studies with data ranging from 1995 to 2017, findings were that, in all cases, High School GPA was as good or a better predictor of college success than SAT score. Over this period, the correlation coefficient for SAT and FYGPA remained consistent, as did the correlation coefficient between HSGPA and FYGPA, with it always being higher for HSGPA than SAT score. Similarly, results for a combined use of SAT score and HSGPA showed consistency.

These studies examined the data, considering other factors that could introduce variability such as test type/format, differences in high school sizes and quality, etc. Bridgeman et al. reviewed SAT, HSGPA, and combined data from 1994-1995 and found that, despite changes in test format and content (e.g. score scale from 1600 to 2400 to 1600, increased length of test, change in question types, allowed use of calculators, etc.), correlations suggested “that neither the content changes nor recentering had much impact on predictive validity” of SAT score for college success (as measured by FYGPA) (3). Similarly, the study by Allensworth and Clark, which looked to determine predictive validity of standardized tests, the ACT in this case, and HSGPA for college success (as measured by college graduation rate) determined that “HSGPAs perform in a strong consistent way across high schools as measures of college readiness, whereas ACT scores do not.” (209)

Furthermore, results indicated that using both sets of data was the best predictor of college success, with the predictability increasing significantly from that of either of the measures used alone. Westrick, et al. indicates that including SAT scores with HSGPA increases predictive power by 15% above HSGPA alone. To elaborate, that study indicates, “Jointly using HSGPA and SAT scores to predict first-year academic performance surpasses the predictive strength of either predictor used alone” (11). It is also important to note that despite a wide range in sample size, from 1,000 to 223,858 students (Table 4), the correlation coefficients for both SAT and HSGPA to FYGPA remained consistent.

In summary, the results of the research conducted on the ETS dataset aligns with the results of these additional studies, indicating that the best predictor of college success is a combination of SAT and HSGPA, followed by HSGPA alone, and finally cumulative SAT score. As colleges examine their admissions policy (and, in some cases, scholarship criteria), it is recommended they consider the results of this and other studies to enable them to make the best decision in predicting college success. Study results suggest that considering both SAT score and HSGPA is the ideal predictor.

In addressing the secondary question regarding any differences based on sex, the analysis showed the variability in first year GPA was remarkably similar between the groups but there was a presence of outliers that reduced each sex’s average first year GPA. Furthermore, the data showed that the mean first year GPA in the female group was higher than that of the male group in this sample and that the increase in mean GPA was even more prominent in the female group compared to the male group. SAT information in the ETS data implied that while the male group in this sample had lower mean high school and first year GPAs than the female group, they had a higher mean cumulative SAT score.

The two-sample t-test validity conditions of independence, normality, and equal variance are all met. The Educational Testing Service used an independent group design to collect the data for SAT, high school GPA, and first year GPA, indicating that we satisfy the validity condition of independence. Furthermore, in the data, there are 516 males and 484 females, meeting the requirement for normality. In reviewing the standard deviations for SAT score distributions, high school GPA distributions and first year GPA distributions for men and women, respectively, we met the requirement for equal variance. That is, the largest standard deviations were not more than twice as large as the smallest standard deviations in each of the comparisons.

Based on the two-sample t-test for FYGPA that resulted in a t-statistic of $t = 34.814$ and a p-value of $p < 2.2 * 10^{-16}$, we have sufficient evidence to suggest that there is a difference between the respective mean first year GPAs of males and females. Furthermore, there is 95% confidence that the true mean first year GPA difference between the sexes is between 0.929 and 1.039. Similar findings were found in the two-sample t-test for SAT score, which had a larger t-statistic of $t = 228.37$ but the same p-value of $p < 2.2 * 10^{-16}$, suggesting that the SAT means between the sexes also differ from each other. This analysis showed that there is 95% confidence that the true mean SAT score difference between the sexes is between 1022.94 and 1.040.67. These trends also followed with High School GPA with substantial amounts of evidence suggesting that each sex's mean High School GPA differs from each other due to the observed t-statistic of $t = 73.53$ and a corresponding p-value of $p < 2.2 * 10^{-16}$. This final analysis indicated that there is 95% confidence that the true mean HSGPA differs between the sexes by between 1.668 and 1.76. All the above tests lead to the conclusion that there is sufficient evidence to suggest that all these metrics differ between the sexes because the two-sample t-test null hypothesis parameter of zero is not included in any of the confidence intervals.

The availability of the data partitioned by sex raised a secondary research question with notable results. This would suggest that there is value in further examination and research of all the data elements (SAT score, HSGPA, FYGPA) to determine if considering data in this way provides further insights into college admissions policies.

Overall, this study provided valuable insights and aligned with the findings of other studies. However, as with any study, there were a few areas that provide an opportunity for improvement. A few limitations of this study were that, unfortunately, the Educational Testing Service did not disclose from which college or region they sampled the 1,000 observations from, making it difficult to generalize based on the analysis of this data alone. Additionally, the ETS data dates to 1996 which leads to some limitations. First, some researchers may claim is too outdated to draw conclusions about the modern relationships between these variables. Second, having more timely and recent data would provide the ability to apply these findings to the population. To improve the study, the ETS could gather more recent data and sample from a variety of colleges in the United States as opposed to one, adding more variability in the results and accounting for confounding variables such as the quality of education.

While the ETS data has some limitations, it is important to note, however, that as stated previously, the studies by Westrick, et al. and Patterson, et al. supplement the analyses from the ETS data. These additional studies were conducted much more recently, used a vastly larger sample size, and sampled from a wide variety of United States colleges. Furthermore, since the results from the tests conducted on the ETS data did not differ significantly from these larger studies, it is plausible that the relationships between High School GPA, cumulative SAT score, and first year GPA have not changed significantly over 21 years. Thus, while it is challenging to make broad generalizations about the United States from the ETS data alone, it is reasonable to do so by examining the similar observational studies conducted by Westrick, et al. and Patterson, et al. Therefore, all these studies suggest that in the United States, examining cumulative SAT score and High School GPA together is the best predictor of college success, followed by examining High School GPA, and then lastly using only the cumulative SAT score.

Works Cited

Allensworth, Elaine M., and Kallie Clark. “High School GPAs and ACT Scores as Predictors of College Completion: Examining Assumptions About Consistency Across High Schools.” Educational Researcher, vol. 49, no. 3, 2020, pp. 198-211.

Bridgeman, Brent, et al. “Predictions of Freshman Grade-Point Average From the Revised and Recentered SAT®I: Reasoning Test.” ETS Research Report Series, vol. 2000, no. 1, 2014, pp.1-16.

"ETS Validation Research: Projects, Papers, and Data." Dartmouth College Department of Mathematics. Accessed 12 Apr. 2023, https://chance.dartmouth.edu/course/Syllabi/Princeton96/ETSValidation.html.

Patterson, Brian F., et al. “Are the Best Scores the Best Scores for Predicting College Success.” Journal of College Admission, vol. 217, 2012, pp. 34-45.

Popham, James W. “All About Accountability / Branded By a Test.” Educational Leadership, vol. 63, no. 7, 2006, pp. 86-87.

Westrick, Paul A., et al. Validity of the SAT® for Predicting First-Year Grades and Retention to the Second Year. The College Board, May 2019, pp. 1-29.