"*Analysis on the performance of regional Sci-Tech students"*

By: Shrey Jain, Veer Khurana, Owen Brake

January 18th, 2019

Table of Contents

1.0 Introduction

Every year at Port Credit Secondary School there are a group of students who pay a $200 fee to part take in a regional Peel program commonly known as “Sci-Tech”. The Sci-Tech program at Port Credit Secondary School is a four-year program that integrates a “HANDS-ON, MINDS-ON” philosophy that provides students with an interactive approach to learning. The merging of science and technology in the classroom enables students to better understand the modern world and prepare students for post-secondary education. The program is enhanced through the “Sci-Tech” curriculum and through participation in activities such as excursions, guest speakers, competitions and clubs.

Thousands of parents have sent their children to specific schools due to the promise of a better education in the fields of science and technology. Individuals are under the impression that attending a regional program will provide them a stronger education, however, this has never been tested before. It costs $200 a year per person to attend the SciTech the program at Port Credit Secondary School, the question is whether all the time, money and effort put into the program is worth it. In order to determine the programs “worth”, the marks achieved by the different groups of students will be analyzed. Marks are an appropriate determinant of academic success as it measures a students ability to perform in the classroom. The question to be tested is “Do Sci-Tech students perform better in their respective Sci-Tech courses compared to those taking similar courses outside of the program?”. Through the immense amount of data and analysis being tested, a sub-question of “Do Sci-Tech students perform at a higher level more consistently?” will be answered. The hypothesis is that Sci-Tech students will perform at a higher level than Non Sci-Tech students in science courses.

2.0 Measures of Central Tendency

2.1 Mean

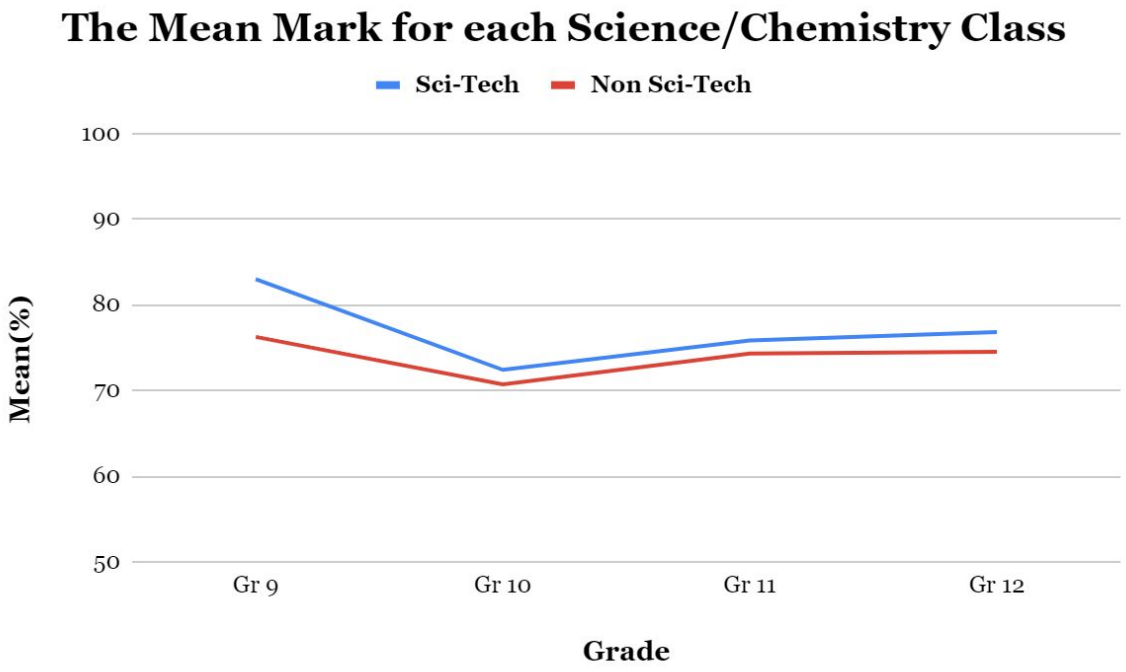

Figure 1. The mean mark for Sci-Tech and Non Sci-Tech Science/Chemistry marks for each grade.

The mean is the average of all the data in the set. The mean of a class or subset of students will approximate how well that subset of students are performing. The graph Figure 1. represents the change in means for each class, across the 4 grades. One surprising piece of information that can be interpreted from this graph is that the mean marks tend to follow the trend for each other in each grade. There appears to be no considerable improvement in Sci-Tech students compared to the Non Sci-Tech students data set, as the non Sci-Tech student's marks decrease so do the Sci-Tech student's marks. Though the two data sets seem to follow the same trend, it appears as though Sci-Tech student's have a small but noticeably higher mean overall then the non Sci-Tech class. The mean accounts for all the data which means it is good at

showing how the entire class is doing compared to other measures of central tendency which are not as affected but outliers and extreme data point values. The fluctuation due to outliers means that the mean variable in this case will not be truly indicative of a standard student in the class, rather it is indicative of the set of data as a whole. With the mean value it becomes misleading to compare two sets of data and say that a normal student in one set performs better than the other set however what can be said is that the class as a whole performs academically better.

2.2 Median

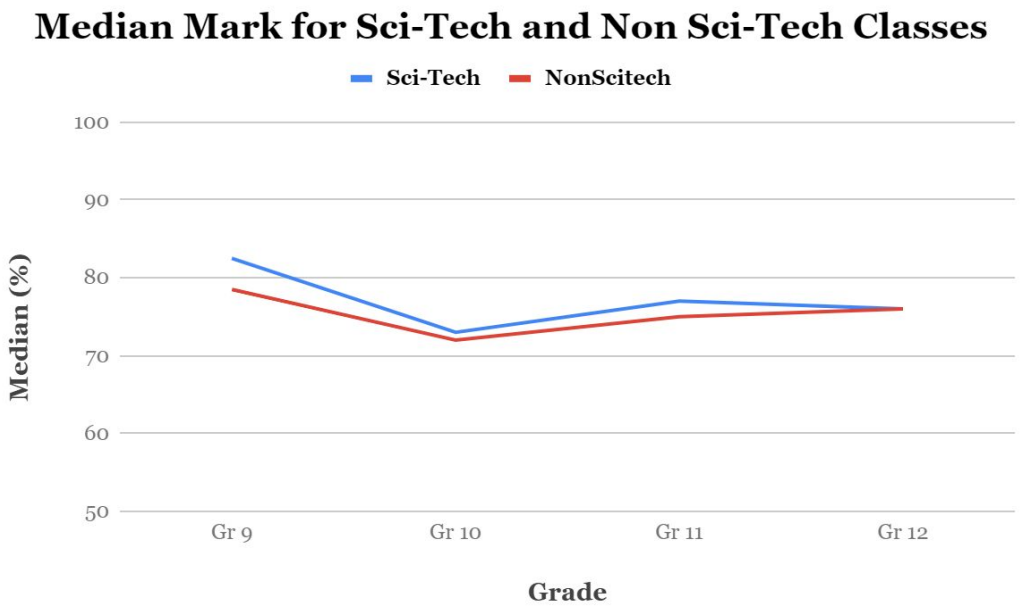

Figure 2. Median mark for Sci-Tech and Non Sci-Tech Science/Chemistry.

The median is the middle value of the data set it is represents the point in which 50% of the data is above and 50% of the data is below this value. As can be seen in Figure 2., the medians tend to be very similar and yet again follow a similar trend. Compared to the graph of the means, in the graph of the medians the difference in marks between those in the Sci-Tech program and those that are not is relatively small. The reason for this change between the mean and the median values is due to the property of medians in which they are not as affected by outliers. Since the medians are closer together it means that in one of the two data sets there are a

significant amount of outliers pulling the mean in an opposite direction to the median. In the case of this data set the median will most likely be a better indicator of how well the standard student is performing as the median will not account for the wild outliers. Thus it can be used to determine whether Sci-Tech students perform better overall simply due to the fact that there are less outliers or is it because Sci-Tech students actually perform academically better. Based on the data from the medians it appears as though the average Sci-Tech student does perform just marginally better than those not in the Sci-Tech program and that there is no significant difference of the trend for the two data sets.

3.0 Measures of Spread

3.1 Standard Deviation

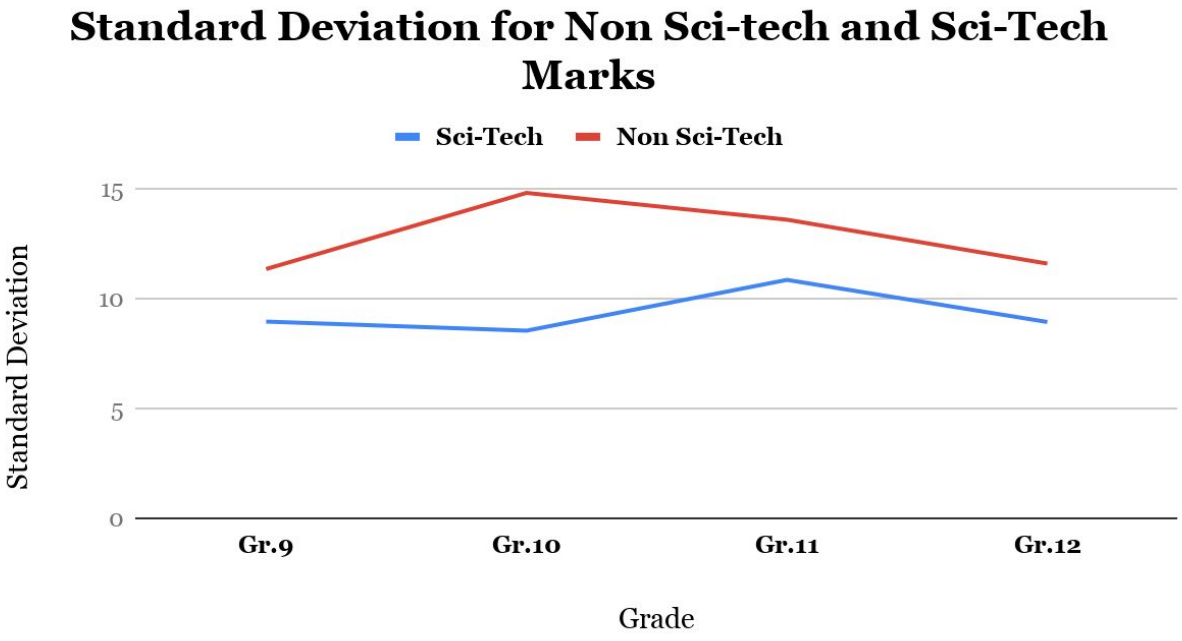

Figure 3. Standard deviation for Non Sci-Tech and Sci-Tech Science/Chemistry Marks

Standard deviation is the measure of how much spread is in the data. A lower standard deviation means there is less of a spread in the data and a higher standard deviation means there is a greater spread in the data. The graph Figure 3. represents the change in the standard deviation for both Sci-Tech and Non Sci-Tech classes throughout each grade. An unexpected piece of information shown in the graph is the increase of the standard deviation from grade 10 to grade 11 for the Sci-Tech Student marks. A possible explanation for this increase is because Grade 11 Sci-Tech Chemistry is a mandatory course for all Sci-Tech students. This results in a wider range of marks because in grade 10 science, certain students might be better at different sciences (Biology/Physics) allowing their final mark to be maintained towards the higher side of marks. However, since Chemistry is a mandatory course, when the student gets into grade 11 chemistry, their mark might be lower compared to other people causing a larger spread throughout the marks. A shocking piece of information to take note of is unlike the Sci-Tech marks, the standard deviation decreases between grade 10 and 11 for the Non Scitech marks. The reason for this decrease in the spread of the marks is because grade 11 Non Sci-Tech Chemistry is an optional course. Therefore students who have less interest and perform academically lower in chemistry will not be taking the course, allowing the marks to have less of a spread and be more consistent. An important trend represented in the graph is that the Sci-Tech standard deviations have all together been lower compared to the Non Sci-Tech standard deviations. Since the Sci-tech standard deviations are lower, it means that the Sci-Tech students overall perform much more consistently in comparison to the Non Sci-Tech students.

3.2 Percentile

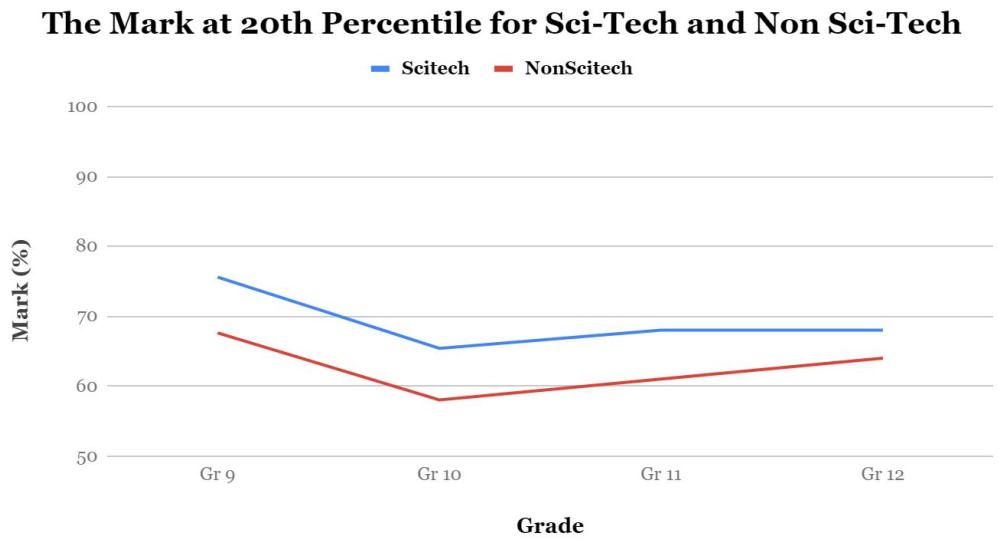

Figure 4. This graph displays the 20th percentile of all students.

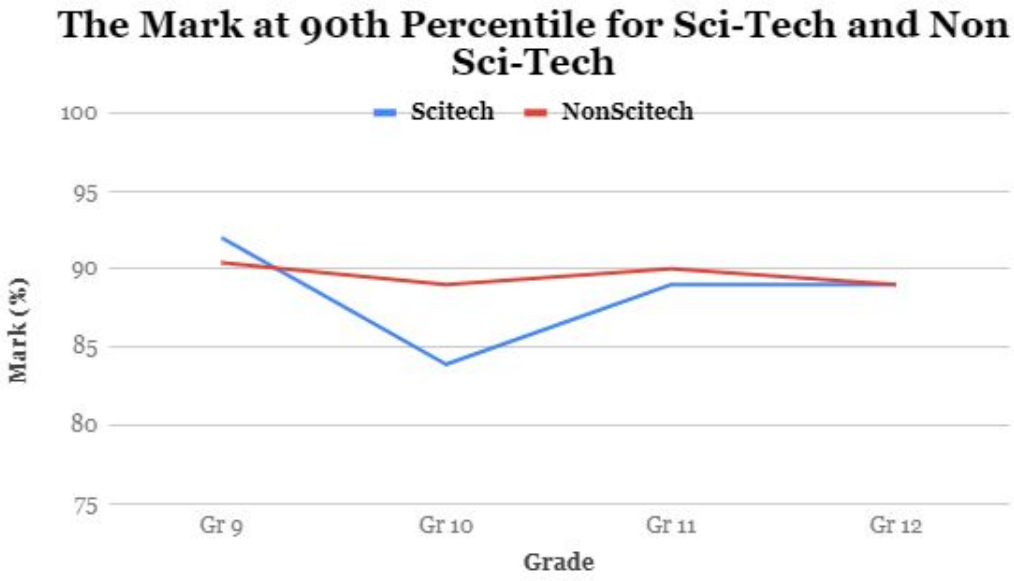

Figure 5. This graph displays the 90th percentile of all students.

A percentile is a measure used to determine the value below which a given percentage of data in a group of observations falls. Comparing percentiles can be very useful as it displays where certain proportions of a class fall in terms of marks. Figure 4. shows the mark for each grade at the 20th Percentile. An important trend to notice is how both the Sci-Tech and Non Sci-Tech graphs have the same shape. However, the Sci-Tech line falls above the Non Sci-Tech line. Sci-Tech students have a higher mark at the the 20th percentile throughout every grade. Since the 20th percentile mark is greater for Sci-Tech, on average Sci-Tech students have a higher concentration of high marks. A possible reason for both graphs to have a similar shape is because each year the difficulty of the course increases or decreases by the same amount for both Sci-Tech Students and Non Sci-Tech students. Figure 5. represents the mark for each grade at the 90th percentile. A shocking piece of information shown is for Grade 10 Sci-Tech, the mark at 90th Percentile is significantly lower. A potential explanation for this drop in marks could be due to Sci-Tech students having to take on a harder course load with more rigorous involvement in Gr.10 science. An unusual piece of information represented by this graph is that overall the Non-Scitech marks at the 90th percentile are very similar to the Sci-tech marks at the 90th percentile. A possible reasoning for this is because in both Sci-tech and Non Sci-Tech there is a large concentration of students who perform very well academically.

3.3 Interquartile Range

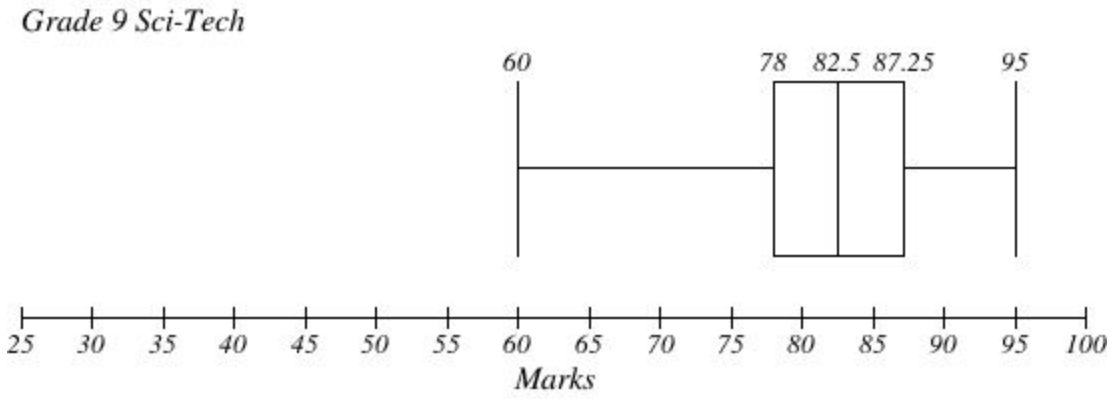

Figure 6. This graph demonstrates the marks for grade 9 Sci-Tech science at the median, 1st and 3rd quartile as well as the maximum and minimum values.

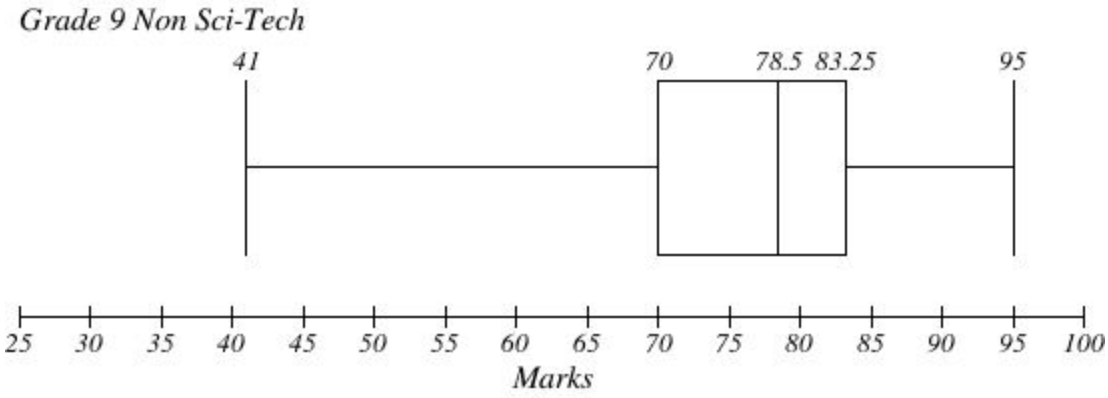

Figure 7. This graph demonstrates the marks for grade 9 Non Sci-Tech science at the median, 1st and 3rd quartile as well as the maximum and minimum values.

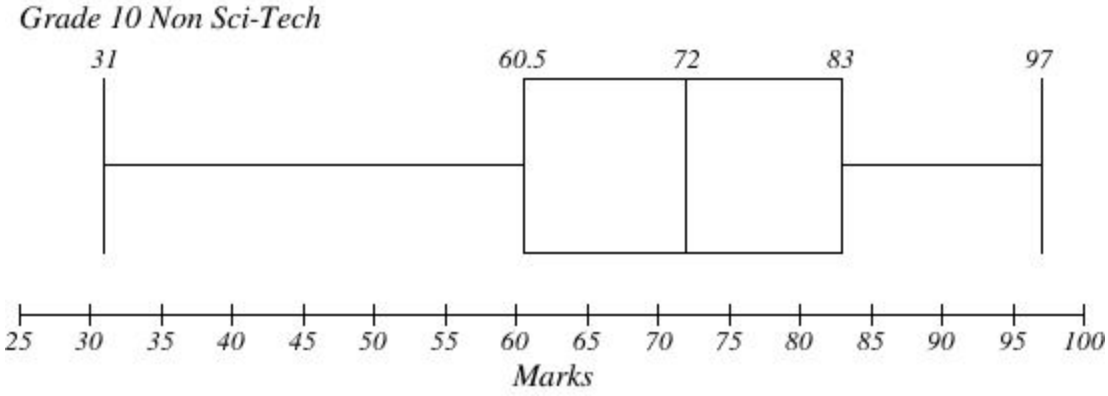

Figure 8. This graph demonstrates the marks for grade 10 Non Sci-Tech science at the median, 1st and 3rd quartile as well as the maximum and minimum values.

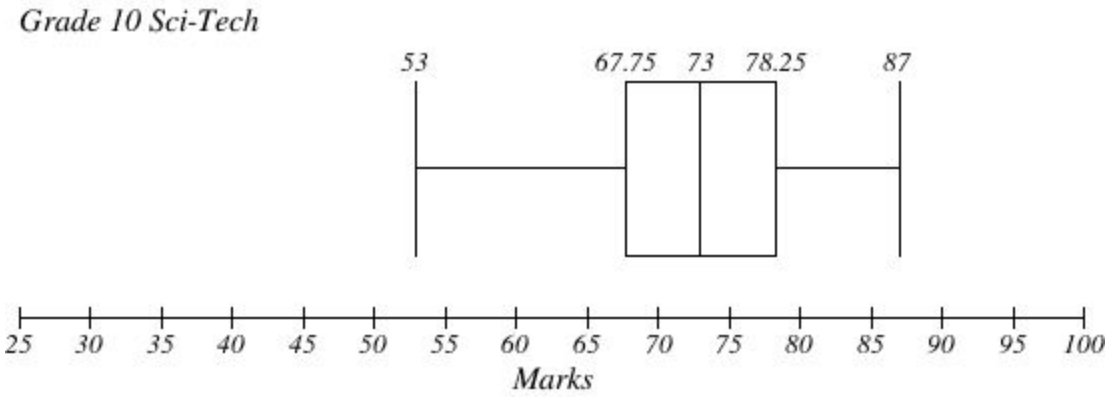

Figure 9. This graph demonstrates the marks for grade 10 Sci-Tech science at the median, 1st and 3rd quartile as well as the maximum and minimum values.

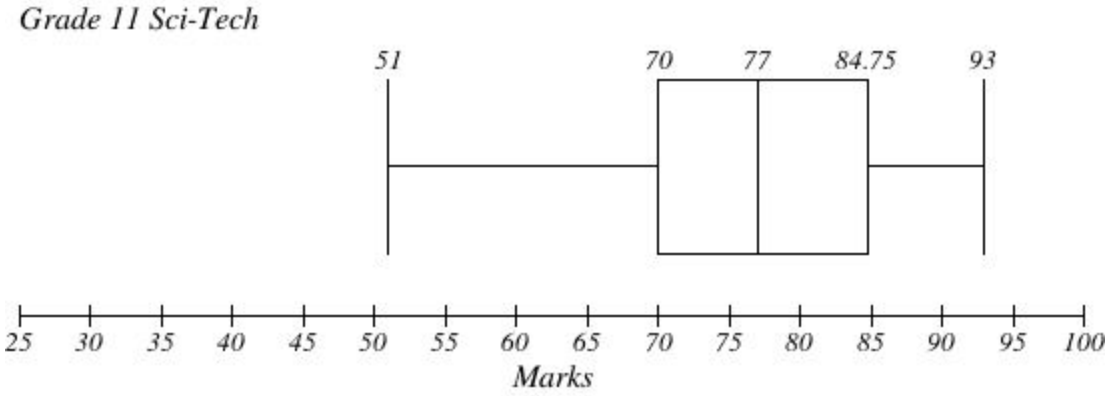

Figure 10. This graph demonstrates the marks for grade 11 Sci-Tech Chemistry at the median, 1st and 3rd quartile as well as the maximum and minimum values.

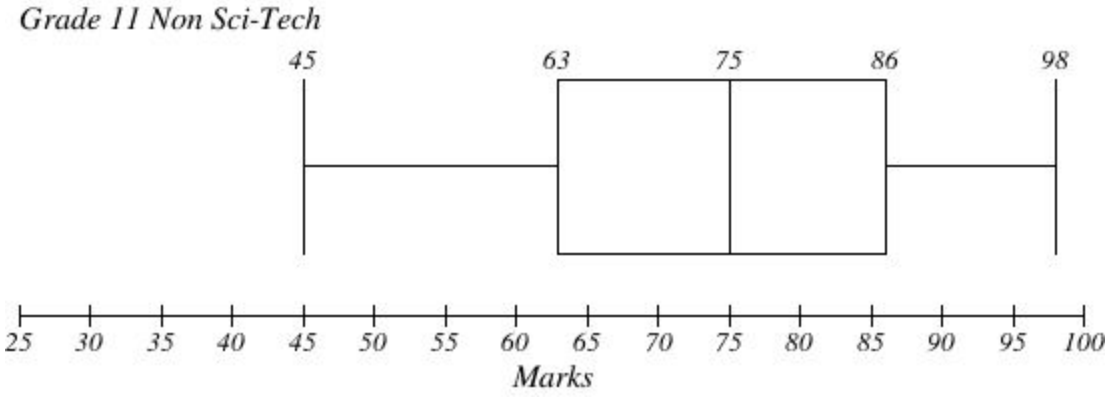

Figure 11. This graph demonstrates the marks for grade 11 Non Sci-Tech Chemistry at the median, 1st and 3rd quartile as well as the maximum and minimum values.

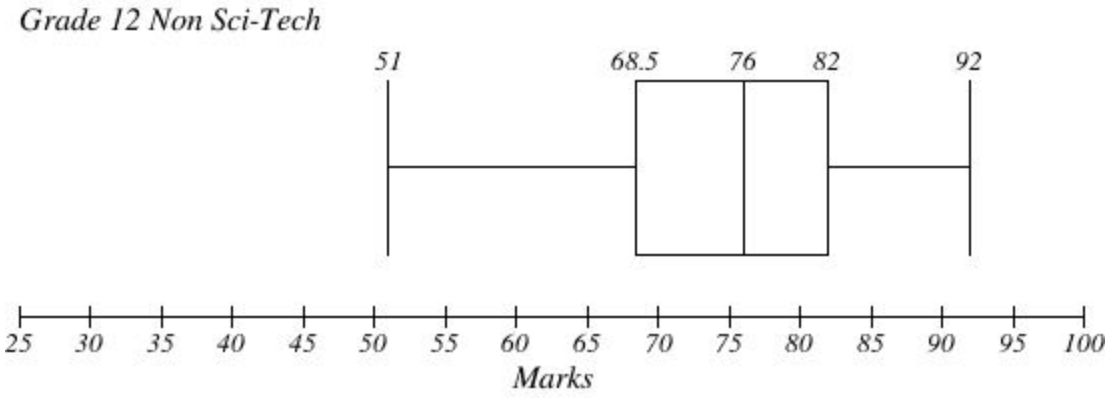

Figure 12. This graph demonstrates the marks for grade 12 Sci-Tech Chemistry at the median, 1st and 3rd quartile as well as the maximum and minimum values.

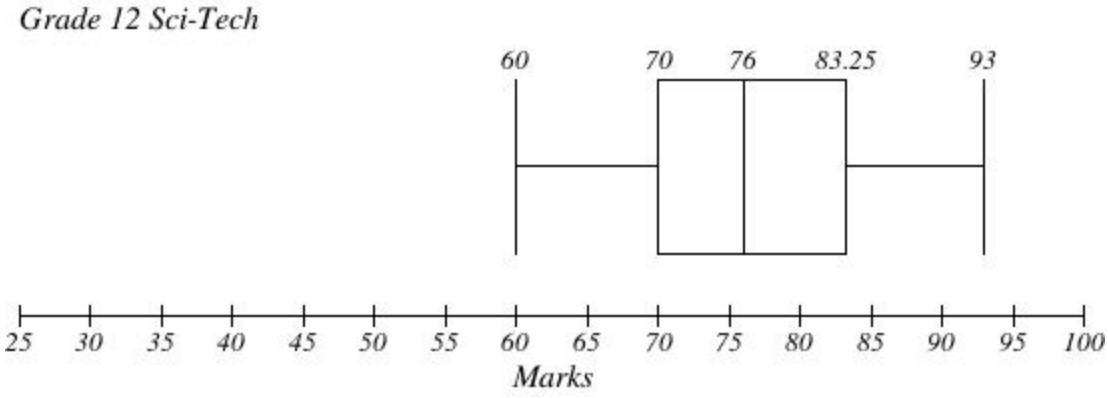

Figure 13. This graph demonstrates the marks for grade 12 Sci-Tech Chemistry at the median, 1st and 3rd quartile as well as the maximum and minimum values.

The interquartile range represents the middle 50% of the data. The following figures (Figure 6, 7, 8, 9, 10, 11, 12,13.) represent the interquartile range for Sci-Tech and Non Sci-tech marks for each grade. For all 4 grades, it is shown that the interquartile range for the Sci-Tech classes are significantly smaller compared to the interquartile range for the Non Sci-Tech classes. Since the interquartile range is significantly smaller for the Sci-Tech marks, it means that the Sci-Tech marks are far more consistent. Another key piece of information represented is the fact that the first quartile for Sci-Tech marks falls comparatively higher compared to the first quartile of Non Sci-Tech marks for every grade. This is means that 50% of the Sci-Tech students fall

above the Non Sci-Tech students in terms of marks. The Interquartile range has a dense concentration of Sci-Tech students that perform at higher level compared to Non Sci-Tech.

3.4 Normal Distribution Analysis

Marks are most often compared by fitting them to a normal distribution. The mark distributions were fit to a normal curve using google sheets. By fitting the marks to a normal distribution it allows the distribution of the marks to be easily visualized providing a clear representation of the consistency as well as how the median values differ. Using the normal distributions to compare the two data sets represents the striking differences between the two data sets that would not be apparent by just representing them with single values such as mean, median and standard deviation.

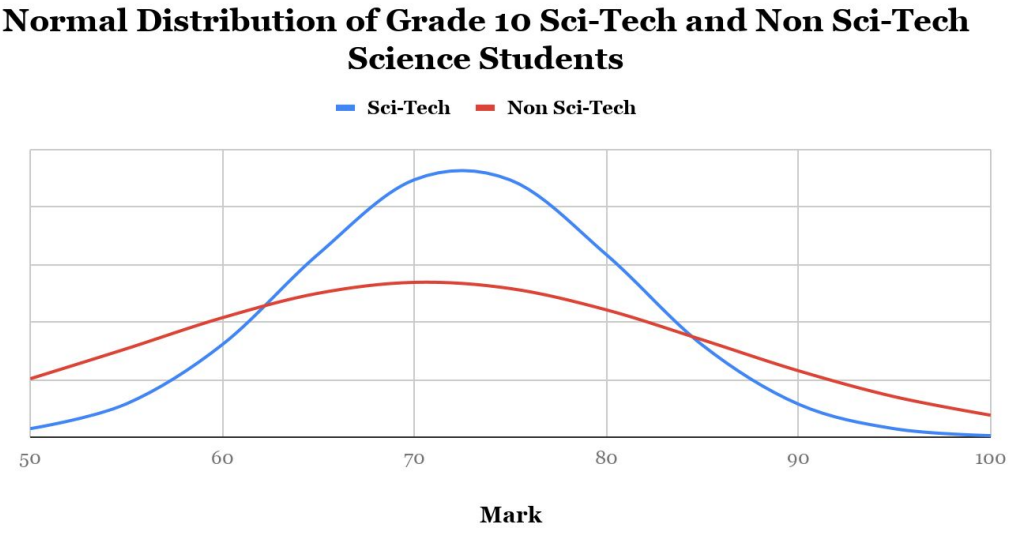

Figure 14. This graph displays the normal distribution of marks for Gr.10 students.

The most striking difference is when analyzing the grade 10 marks for Non Sci-Tech and Sci-Tech marks in Figure 13., as can be seen in those graphs, the Sci-Tech marks have a very high concentration in a small range and there are a small amount of marks distributed out on the tail ends of the distribution. The Non Sci-Tech marks however do not represent a perfect normal distribution rather it appears almost flat, this is indicative that the class is incredibly inconsistent and that the marks are not concentrated in any location. Even though the Grade 10 Non Sci-Tech

marks appear incredibly inconsistent it should be of note that on the tail ends of the distribution there are a larger concentration of marks then in Sci-Tech meaning that the best Non Sci-Tech students do better than the best Sci-Tech students however the worse students in Non Sci-Tech also do worse then the Sci-Tech students.

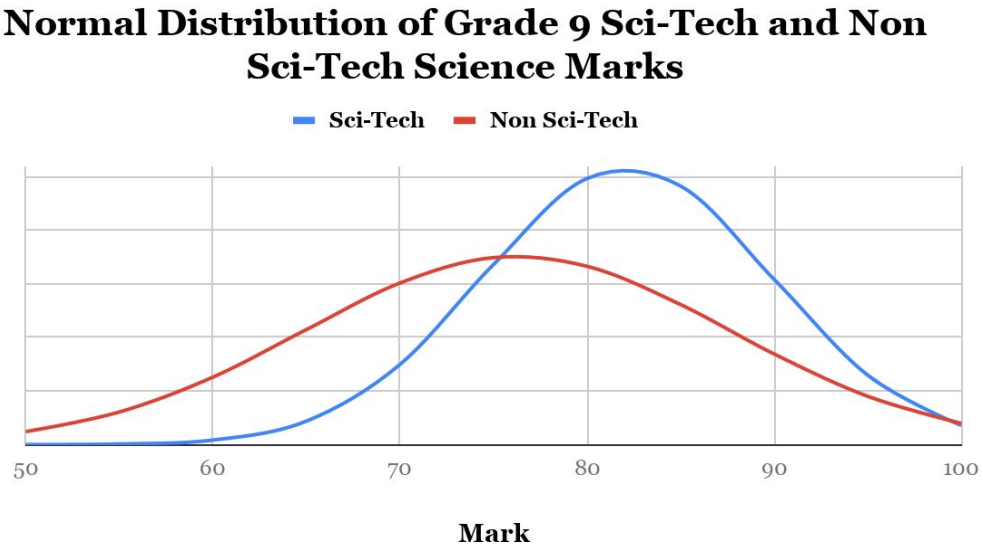

Figure 15. This graph displays the normal distribution of marks for Gr.9 students.

Another very important graph is that of the grade 9 marks represented in Figure 14. When viewing the grade 9 marks it gives a strong indication that the hypothesis is correct, all the indicators in this graph show that Sci-Tech students perform at a higher level in comparison to Non Sci-Tech students. The normal distribution for grade 9s show that the peak of the curve for Sci-Tech students is significantly higher than that of Non Sci-Tech students meaning that the former tend to perform better then the latter. The Sci-Tech students also have a tighter curve representing that they are more consistently performing higher then the Non-SciTech students.

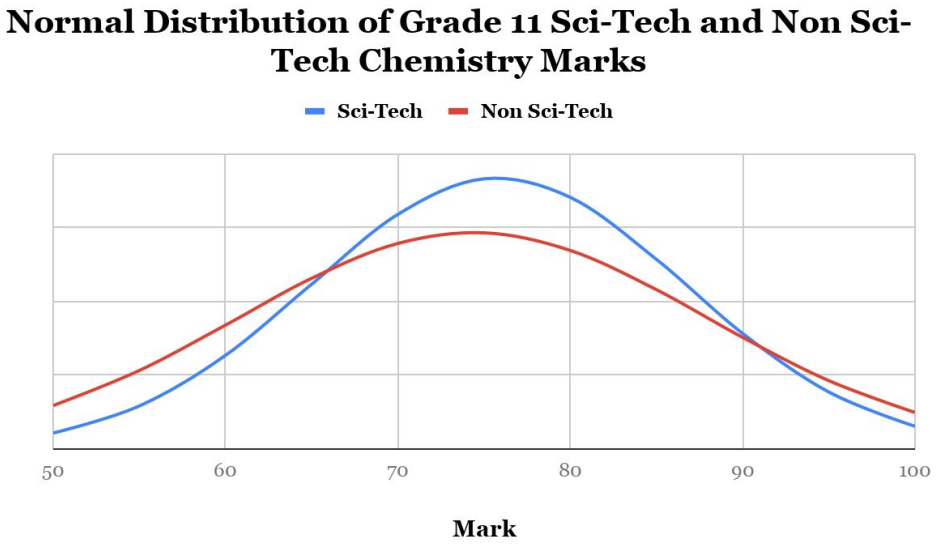

Figure 16. This graph displays the normal distribution of marks for Gr.11 students.

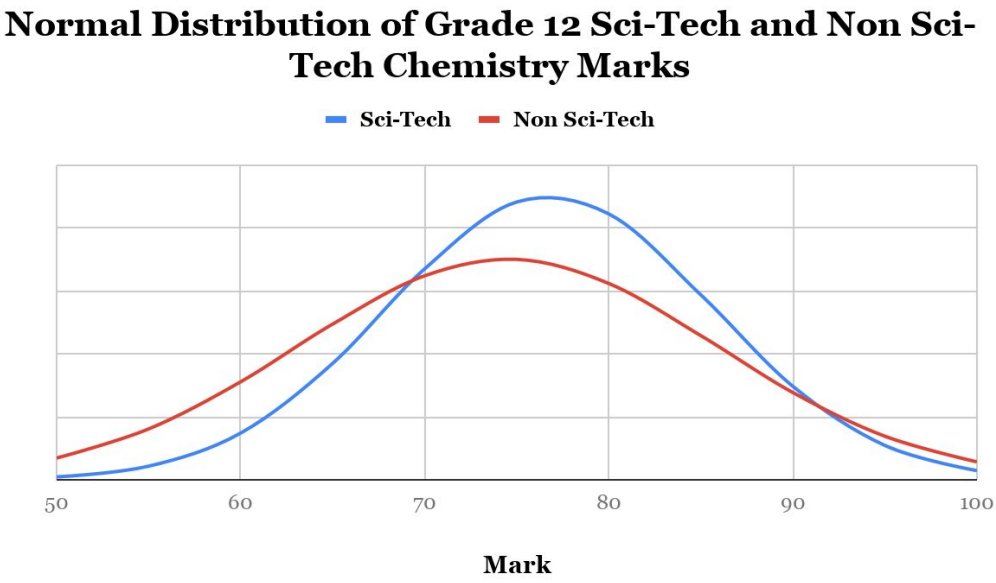

Figure 17. This graph displays the normal distribution of marks for Gr.12 students.

Finally there are the grade 11 and grade 12 graphs represented in Figure 15 and 16. Both these graphs are very similar and appear to show a similar trend. Yet again the Sci-Tech students have a higher peak and steeper curve representing that the Sci-Tech students are more consistent and that most of the data points are centralized around one area. The Sci-Tech curve peak is also slightly placed to the right of the Non Sci-Tech curve representing they tend to perform better.

All these trends have already been identified however an important point about these two graphs is that they both considerably show the Non Sci-Tech students and the Sci-Tech students being significantly closer to each then the first two graphs. In the graphs representing the grade 9 and grade 10 marks the difference between the two curves is stark however it appears as the grade increases the two curves in each graph become more similar.

3.5 Hypothesis Test

The hypothesis test is used to determine whether or not Sci-Tech students academically outperform the mainstream students is constituted as statistically significant and not caused by sampling error.

The hypothesis test was conducted at a significance level of 1%. Let H_o represent the null hypothesis and H_1 represent the alpha hypothesis. H_A represents the performance of Sci-Tech students and H_B represent the performance of mainstream students.

$H_o = H_A = H_B$
$H_1 = H_A > H_B$

The hypothesis test formed is a one-sided hypothesis. A one-sided hypothesis claims that the parameter of H_A is greater than the parameter of H_B.

Data Values:

Variance of Sci-Tech Students = $\sigma_A^2 = 72.65$	Variance of mainstream students= $\sigma_B^2 = 149.0104$
$\bar{X}_A = 76.495$	$\bar{X}_B = 73.935$
$n_A = 209$	$n_B = 229$

$$z = \frac{76.495 - 73.935}{\sqrt{\frac{72.65}{209} + \frac{149.01}{229}}} \qquad z = \frac{\overline{x_A} - \overline{x_B}}{\sqrt{\left(\frac{\sigma_A^2}{nA} + \frac{\sigma_B^2}{nB}\right)}}$$

z= 2.56

Therefore, the data collected and the results shown are not a result of chance or sampling error. This in turn indicates that regardless of the cohort selected, the data would remain true with a minimal amount of change. The P of 0.9948 indicates that the null hypothesis is going to be rejected and the alpha hypothesis must be accepted with confidence. The P value that was calculated falls within the critical range of the normal distribution.

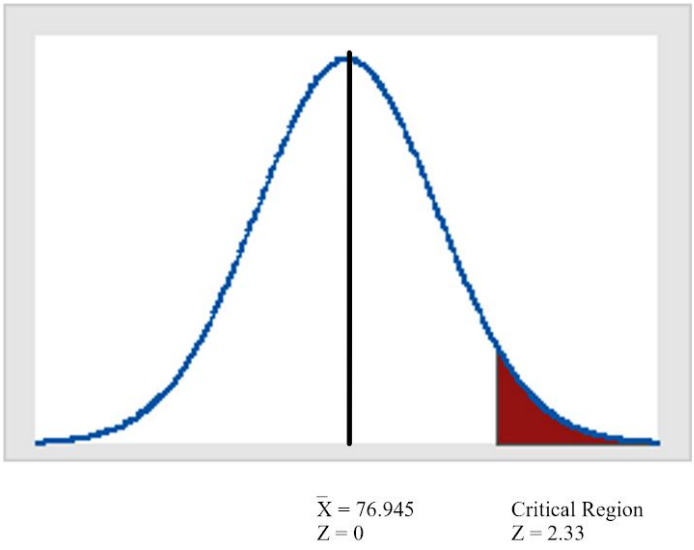

Figure 18. This diagram displays the normal distribution of students marks and specifically indicates the critical region.

Overall, this proves that there is a difference between the performance of Sci-Tech students and mainstream students, in that the Sci-Tech students outperform the regular students. However, let it be known that many variables were not taken into account in this hypothesis test.

Specifically, all of the extraneous variables outline further in the report cannot be accounted for in this hypothesis test. Thus, it cannot be proven that it is statistically significant.

4.0 Correlation Coefficient, r

In order to determine the relationship between Sci-Tech and Non Sci-Tech students academic performance, a correlation between the median of each group of students was graphed and analyzed in Figure 14. . In the figure provided the dependent variable is the Sci-Tech classes' median marks for that grade and the independent variable is the Non Sci-Tech class median marks for that same grade. This method of plotting the median marks provides an analysis method which is independent of certain extraneous variables the difficulty of each class as each point plotted only represents the difference between the Sci-Tech and Non Sci-Tech students taking nearly identical courses rather then comparing across different years. A linear regression was then performed on the data set so as to determine whether if there is a correlation in the data and if so what that correlation is. The correlation is used to determine whether or not there is a relationship between how well Sci-Tech students perform and how well Non Sci-Tech students perform. The reason why the median was tested was to ensure that the data was not skewed by any outliers in the data sets and to ensure the graph was indicative of the performance of a standard student.

The line of best fit regression equation was calculated as $y = 0.637x + 26.23$ for Figure 14..This equation gives us two variables an m value of 0.637 and a b value of 26.23, the most important value in this case is 0.637. An m value in a linear equation can be calculated as $m = \frac{\Delta y}{\Delta x}$ which means that the m values is reflective of how much the dependent variable increases in comparison to how much the independent variable increases. Therefore since the m value is between 0 and 1 then the independent variable increases at a greater rate then the dependent variable. In the case of the graph provided the independent variable is the Sci-Tech marks and the dependent variable is the Non Sci-Tech marks, thus the Sci-Tech marks increase at a greater rate then the Non Sci-Tech marks.

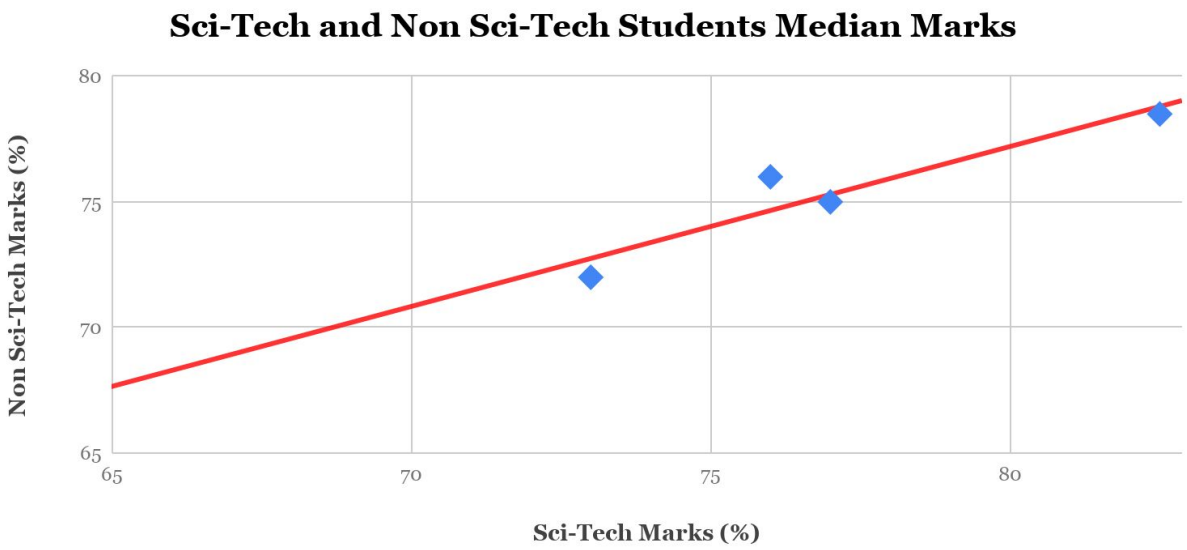

Figure 19. This chart displays the correlation between the median marks of Sci-Tech and Non Sci-Tech students. r

There are two main variables that need to be assessed when analyzing the strength of a correlation between Sci-Tech students and Non Sci-Tech students; r and r^2. The r and r^2 determine the strength of the relationship between the data points on the scatter plot and the linear regression line. The greater the r and r^2 value, the greater the strength and meaning of those points. The r value for this linear regression is 0.939, the r value in this case represents how strong the Sci-Tech and Non Sci-Tech variables are correlated while the r^2 which is 0.883 is how well the data fits the regression. Both of these variables indicate a strong correlation however it is the r^2 value which tells us that the data matches the regression to a statistically significant amount and that do perform better than Non Sci-Tech students.

5.0 Linear Regression and Interpolation/Extrapolation

The other purpose for performing a linear regression is for the purposes of extrapolation and interpolation. By performing an interpolation or extrapolation of the data, future marks can be predicted and the marks of Non Sci-Tech classes can be used to predict what a Sci-Tech classes' median would be.

An example of an extrapolation would be if a Non Sci-Tech classes' median is 85 and a school is looking to implement a Sci-Tech version of that class they may predict the median by solving for the Sci-Tech median. Once the value of the median 85 is plugged in the equation $85 = 0.637x + 26.23$ is given and is solved to result in a median mark of 92.26 for the Sci-Tech class. There is however a large error when solving for this extrapolation, most likely if another Sci-Tech course was added and the Non Sci-Tech course has a median of 85 the median of the Sci-Tech course would not be 92.26. The error is that there is such a small set of data that a proper/accurate extrapolation would be nigh impossible as there is no data to indicate that this trend will continue for an extended period of time, whether the regression will curve or change at different levels. Having four points on a scatter plot, performing a regression and then expecting that regression to hold true for the rest of the plot is a fallacy. To prove this regression is not perfect one can simply say that the Non Sci-Tech course median is a 90 and solve $90 = 0.637 + 26.23$ to get a Sci-Tech median of 100.11, which is impossible.

An example of an interpolation would be if in first semester there is a median of 75 for the Sci-Tech class and the staff are attempting to predict the median for the Non Sci-Tech class in second semester. The staff can interpolate by solving the equation $y = 0.637(75) + 26.23$ to get a Non Sci-Tech class median of 74.00. This interpolation is significantly more accurate then the extrapolation performed earlier as it has more relevant data to compare to, there is a high concentration of data around the median 75 and thus the regression will fit that data better. This interpolation may be an appropriate representation of what the Sci-Tech class may achieve based

on previous data yet again it fails to account for several key factors and variables which will cause the data to fluctuate significantly. Overall an interpolation would be more accurate at predicting future data points then an extrapolation and any method of prediction should be performed with caution, with the knowledge that the trend if not perfectly accurate.

6.0 Analysis

The initial question posed was “Do Sci-Tech students perform at a higher level in science courses compared to Non Sci-Tech students?”. As proved through the analysis, Sci-Tech students do in fact perform at a higher level, but this is largely due to extraneous variables and is not caused by the relationship between Sci-Tech students and Non Sci-Tech students. The relationship between these two variables is a presumed relationship. A presumed relationship is one in which a correlation does not seem to be accidental and there is no cause and effect or common cause relationship between the two variables. When taking a look at the relationship between Sci-Tech and Non Sci-Tech, its shown there is a correlation, however, after analysis, its realized that there is no cause and effect. There are many variables that need to be taken into account when analyzing the outcomes.

Throughout the analysis many different variables were used to analyze the data between Sci-Tech students and Non Sci-Tech students. These variables include: the grade of the students; the type of student (Sci-Tech or Non Sci-Tech); mean, median, percentile, variance and standard of marks obtained by students.

6.0.1 Grades

The specific grade in which students at Port Credit Secondary School are currently enrolled in was analyzed to display the change that occurs over a four year period. Grades is a categorical variable as it is a qualitative value to assign students to a specific group. The grade in which students were in was used in many different analysis's. When analyzing figures 1-12 the Grades were used on the x-axis to display the change in students mark over the course of four

years. Therefore, grades are an independent variable as the marks that students obtain are dependent on the grade the student is in.

6.0.2 Sci-Tech and Non Sci-Tech

The differentiation between Sci-Tech students and Non Sci-Tech students was key in determining the performance difference in these groups of students. These two variables categorical variables as it is a qualitative value that it is used to identify a group of students. In figures 1-12 both Sci-Tech and Non Sci-Tech were independent variables. The mark that a student obtained was dependent on the program the student was enrolled in. In Figure 14 , both Sci-Tech and Non Sci-Tech marks were graphed with Sci-Tech on the x axis and Non Sci-Tech on the y axis. However, this does not indicate that Sci-Tech is an independent variable. In this specific case, it did not matter where each group was placed, but rather analyzing the regression model to show which one out performed the other.

6.0.3 Mean, Median, Percentile, Variance, Standard Deviation in Marks:

The mean, median, percentile, variance and standard deviation were key variables used when analyzing the performance comparison of students. All of these values are continuous variables as the are a numeric value represented by a percentage. The percentage obtained by these students are continuous variables because they represent a measurement. All of these variables are also dependent variables; they are dependent on the grade and type of program students are enrolled in.

6.1 Extraneous Variables

Additionally, there is abundant amount of extraneous variables that need to be taken into account when analyzing the data. These variables include: a students academic background before entering high school; various concentrations of academically ambitious students in the

same classroom; chemistry mis-representing a students ability to perform in the sciences; and style of teaching in the classroom.

6.1.1 Students background before entering high school

Before entering high school, students can come from any middle school. These middle schools could be private or public middle schools. These middle schools could be located in different areas of the country, where the education system differs. Additionally, the middle school education system differs greatly depending on the school that you attend. Only 31% of Canadian school have specialized teachers teaching mathematics and science. The overall funding being put into the middle school education system is a large variable in creating a falsified preparation for these students prior to entering high school.

Currently, majority of the Gr.9's entering Port Credit Secondary school come from all different areas of the Peel region and abroad. The school is a huge factor in preparing these students for high school. If a student attended an enriched private or public middle school, there is a high chance that the student will more academically inclined to participate in rigorous or challenging content in high school. Additionally, a huge factor that is associated with a student's middle school is the ***socioeconomic status*** of the family. Students that have families that are able to support them are at an advantage before entering high school. These students will not have to worry about working throughout high school and comprising education for money. Additionally, families of higher socioeconomic status are generally well educated, allowing the family to pass down their education and assist their students in their academic pursuits. Thus, these students are going to receive an immense amount of support from their families enabling them to have a competitive advantage.

6.1.2 Various concentrations of academically ambitious students

Students entering the Sci-Tech program are responsible for filling in an application in which the marks of a student are required. Over 500 students apply each year. 240 of these 500 are selected to attend a screening process where a final 135 are selected. The filtering of this

process alone is automatically going to differentiate students entering the Sci-Tech program versus the Non Sci-Tech program. Students who achieve strong academic achievement in middle school are selected to attend this program. Thus, the program is going to be concentrated with a group of students who are already more academically capable than the average student.

Additionally, this concentrated group of students are also academically ambitious for the future. Students come to Port Credit Secondary School just for the Sci-Tech program from all areas of the GTA. There are a large group of students which commute over 3 hours a day just to be able to attend the Sci-Tech program, so they will not be obligated to attend their local secondary school. The ambition and willingness that a student must have to be able to dedicate this amount of time into attending a more specialized high school is indicative of their ambition for the Sci-Tech program and its specialized courses..

The high concentration of academically ambitious student in the same classroom is a major factor in enhancing the performance of Sci-Tech students versus the regular Non Sci-Tech student. Being immersed in a classroom setting in which students are competing to be at the top of their class provokes every student to aim for high academic achievement. This immense amount of competition is associated with an increase in work ethic. Students feel the need and desire to be at the top, and as competition grows, it requires more work and more studying enabling students to perform better in the classroom.

In contrast, looking at a group of students in a Non Sci-Tech class where there is a large standard deviation in academically performing students, as represented in Figure 3., creating inconsistency within the classroom. This in turn forces the teacher to teach material more than once and only touch the main concepts and go into as much depth that would normally be reached. Overall, the lack of students who are ambitious in the classroom in the Non Sci-Tech program negatively impacts the quality of education.

6.1.3 Chemistry is a misrepresentation of science abilities

Currently, the Sci-Tech program mandates that students enroll in a variety of science courses including: Gr.9 Science; Gr.10 Science; Gr.11 Chemistry; and Gr.12 Chemistry. However, not every Sci-Tech student is interested in pursuing anything related to chemistry in the future. The only reason why these students take the course is to complete the required courses for the academic program. This can be a a huge factor in skewing the relationship between our data because students motivation and work ethic in these courses varies in upper years. Chemistry does not represent a students ability to perform in all science courses. Yet, the only Sci-Tech science in the senior years of high school that Port Credit Secondary School currently offers is Chemistry. Allowing students to pursue the science courses which they are passionate about and not forcing them to take courses which are unnecessary for their post-secondary goals may incite a statistical increase in the marks obtained by students.

6.1.4 Style of teaching and impact of teachers

The Sci-Tech program advertises that the teachers teaching Sci-Tech courses are "selected on the basis of their diverse teaching and work related experiences and are committed to providing students with a relevant, practical and stimulating learning environment". Port Credit Secondary School has an immense amount of qualified teachers in their respective fields. The Sci-Tech department specifically has teachers that specialize in the courses they are teaching, allowing for the classroom to be more engaging and interactive with questions and depth of learning. The teachers differ from Sci-Tech class to Non Sci-Tech class, and sometimes overlap.

After interviewing one of the lead Sci-Tech Chemistry teachers, Alka Saxena, the different methods of teaching between regular Sci-Tech science courses and Non Sci-Tech science courses was gathered. Ms. Saxena claimed that when she teaches Sci-Tech courses she

feels that the “the class challenges [her] and is competitive and everyone is at a similar academic level”. This is important because teachers to not have to waste time on simple concepts and be able to teach the class more advanced topics. She added “in Sci-Tech classes, [she] is given the opportunity to challenge students through hands-on projects, giving students the opportunity to collaborate with one another. [She] found that as students get into upper years, the develop stronger work habits, especially Sci-Tech students.” Ms. Saxena also claimed that Sci-Tech students and Non-Sci Tech students get tested on the same material for tests, assignments and examinations. This is important to note, as Sci-Tech students have the opportunity to learn more advanced topics, but still only get tested on the Ontario regulated curriculum. Sci-Tech students practice complex problems to prepare them for future grades, or post-secondary education, but are only tested on the regulated information, thus allowing them to perform at a higher level. Overall, the style of teaching differs based on the class and the teachers experience has an immense impact on the way that a student performs academically.

6.2 Summary

Overall, the mass amount of variables that need to be taken into account when analyzing the data indicate that these two variables are connected via a presumed relationships.

7.0 Conclusion

7.1 Bias

Although the hypothesis test showed there was a minimal chance of sampling error, there was bias in the data. The type of bias in the data collected was sampling bias. Sampling bias is when the chosen sample does not represent the entire population. This is true for a multitude of reasons.

Firstly, the data that was gathered was from the 2019 school year at midterm marks. This specific cohort of students could have potentially been a unique set of students where trend

between Sci-Tech and Non Sci-Tech was present. However, this may have not been true for all groups of students. Additionally, the only marks that were analyzed were the midterm marks. By midterms in high school, students have only accomplished a minimal amount of the entire curriculums content and have the ability to raise their mark by a significant amount. Thus, analyzing just the midterm marks is not sufficient. These two issues of sampling error could be solved through looking at the final marks of students over the past 4 years so there would be a lot more data to confirm that the results that were found are true with a greater confidence level.

Additionally, the only comparison between students in both programs was done through the comparison of the mark obtained in science or chemistry students. The relationship between Sci-Tech students in science courses and Non Sci-Tech students in science courses should display a difference, as these students came to Port Credit to specialize in science and technology. What may be a further question to ask and analyze, would be do Sci-Tech students perform higher than Non Sci-Tech students in non science and technology courses.

The initial question of do Sci-Tech students perform at a higher level in science courses in comparison to Non Sci-Tech students is only answered with the data obtained by students at Port Credit Secondary School. There are multiple Peel Sci-Tech programs and STEM based schools globally. The results gathered through this report should not be used to generalize the performance of STEM students in general as the data of marks from different school were not taken into account.

Overall, there are many different variables that induced sampling bias into this research study. In the future, taking into these variables would lead to a stronger study that would be able to prove the relationship of STEM students in comparison to mainstream students with a greater confidence interval.

7.2 Summary

Based on the findings in this report a few conclusions can be made about the data as well as the correlation between Non Sci-Tech and Sci-Tech students' marks. The primary conclusion that can be made is that the hypothesis stated in the introduction cannot be proven with the given data set. The current data set has failed to prove that there is any significant statistical link between Sci-Tech students and higher academic performance. The only reasonable conclusion to the hypothesis, given the current data set is that Sci-Tech students do not perform academically better than Non Sci-Tech students and if so not by a statistically significant amount. Any seemingly clear relationship between Sci-Tech students and academic performance in the report can be written off as a presumed relationship. All of the extraneous variables need to be taken into account when developing a relationship between these two variables. Through the compilation and analysis of the data it was found that while Sci-Tech students do appear to have an academic advantage in the earlier grades such as grade 9 and grade 10, this advantage smooths out and the data sets become more similar in the senior years of high school. Thus, by the senior years of high school, the academic performance between these two groups of students are about the same. Over the course of the four years of high school, the Non Sci-Tech group is consistent with its grades. However, the Sci-Tech group of students has a decline in marks allowing it to reach equilibrium with all other students. It can therefore be stated that it does not matter whether or not a student is in the Sci-Tech program pursuing science courses, as it was found that eventually both Non Sci-Tech and Sci-Tech students are able to achieve the same grades. However, let it be known that the extraneous variables such as the community, skills developed, and opportunities for Sci-Tech students differ to others. In the big picture, a STEM based program such as Sci-Tech does not enhance the academic performance of a standard student in science courses.

8.0 Works Cited

"Australian Bureau of Statistics." *Statistical Language - Measures of Central Tendency*, www.abs.gov.au/websitedbs/a3121120.nsf/home/statistical language - measures of spread.

McKenzie, Helen. "Education in Canada: Current Issues." *Education in Canada: Current Issues*, Political and Social Affairs Canada, May 1994, publications.gc.ca/Collection-R/LoPBdP/BP/bp386-e.htm.

Meisel, Roland W. *McGraw-Hill Ryerson Mathematics of Data Management*. McGraw-Hill Ryerson, 2003.

"Peel Board Student Information System." *Peel Board* , Student Information System, 15 Nov. 2018.

"The Normal Distribution." *Categorical Data*, www.stat.yale.edu/Courses/1997-98/101/normal.htm.