# STATISTICAL ANALYSIS OF A CATHOLIC CONFIRMATION PROGRAM

**James Fulford '17***
**Undergraduate Student, B.A. Program in Mathematics, Rivier University**

**Abstract**

*The dataset analyzed consists of 568 high school sophomores engaged in the confirmation preparation program at St. Elizabeth Seton Church from 2013 to 2016. Analysis was performed on global and subset data to determine potential relationships between confirmation rates, student volunteering, and student absenteeism. Results of particular interest include summer courses having significantly higher volunteering and lower absenteeism, at the expense of significantly lower confirmation rates. Also, the lack of a correlation between student volunteering and student absenteeism is of particular interest. On the whole, however, the confirmation rate is high.*

## INTRODUCTION

### Motivation and Objectives

According to dynamiccatholic.com, more than one million young Catholics prepare for Confirmation every year. However, by the age of 21, approximately 70% of these young Catholics stop attending Mass altogether, rising to 85% by the age of 30. Before this sudden drop in attendance, most (90%) of these young Catholics attend a confirmation preparation program. Furthermore, according to religionnews.com, recent studies have found one out of every four Catholic parents do not find importance in their children being confirmed. These statistics seem to indicate that young Catholics are not engaging in their faith in the United States.

The objective of this study is to establish whether or not there is a statistical relationship between confirmation rates, student absenteeism, and student volunteering at St. Elizabeth Seton Church in Bedford, New Hampshire. The findings of this report could potentially be used to identify best practices and critical subsets in Catholic confirmation education.

The researcher is a graduating Mathematics undergraduate student at a Catholic university who has completed a Catholic confirmation program, despite strong personal reservations at the time. Mainly through parental guidance, the researcher went on to enthusiastically receive the Sacrament, teach the program to a younger cohort, and still attend weekly services.

### Program Description

The majority of students (n = 468) enrolled in a year-round format, which met on a monthly basis on Sunday afternoons for 7 sessions per school year, delivered over two school years. In 2014, enrollment in a summer format began (n = 100). Summer students met on a daily basis for a week and a half in late June, generally from 10 am to 3 pm. As part of the confirmation program, both summer and year-round students were strongly encouraged to volunteer at least five times for the parish, soup kitchens, nursing homes, or other community institutions.



ISSN 1559-9388 (online version), ISSN 1559-9396 (CD-ROM version).

In both formats, students engaged in large group and small group activities. The small group classrooms were led by teaching volunteers (generally parents of confirmation program students) using the same curricular resources (activities, textbooks, lesson plans, etc.). When in large groups, students watched videos, listened to guest speakers, and received lectures. Snacks and socialization opportunities were scheduled between lessons.

Data was retrieved directly from the church's director of the Confirmation program, Ms. Carrie Soucy. Excel spreadsheets and photocopies of volunteering records were provided, in addition to the hard-copy program attendance records and a printed list of students who attended the final ceremony.

Boolean data was established on whether students were confirmed or not. A count of the number of volunteering events attended was established for each student. Lastly, a measure of absenteeism was established for each student based on how many sessions of the program they missed fully or partially.

## DEFINITIONS

### Cohorts

Some subsets of the students are of particular interest to analysis. For instance, the subset of students who attended the summer format may behave different than year-round format students. The researcher has identified 11 subsets of the dataset of particular interest. These subsets, referred to throughout the paper as cohorts, are listed and described in Table 1 (see Appendix A).

### Variables

#### *Confirmation Rates*

A student is considered confirmed if they attended the St. Elizabeth Seton confirmation ceremony the same year they finished the preparation program. Some students delayed confirmation for a year or more (often to be confirmed alongside a sibling), while others or were confirmed in another parish. The reason delayed confirmation is not counted in this study is because of the 2016 students who have not had the opportunity to get confirmed the following year.

In general, the program had unusually high confirmation rates, at 93.3% over the years studied. A list was provided by the program coordinator of who attended the final confirmation program for each year. If a student was not on this list, it was presumed that the student was not confirmed at St. Elizabeth Seton that year.

Students were only compared to the list of ceremony attendees if they committed to completing the second half of their confirmation program. To the knowledge of this researcher, only 6 students dropped out of the program prior to the second half, which are ignored in this study.

#### *Student Volunteerism*

Volunteering events mostly consisted of visiting nursing homes, committing hours at food pantries/soup kitchens, and/or helping with parish events (dinners, youth programs, etc.). Longer volunteering events occasionally counted for more.

Database reports and paper photocopies were provided by the program coordinator for most students in the program. Each entry included the name and date of the volunteering event, however this study only considers the total number of entries for each student. For each entry, sufficient evidence that the student participated was submitted by the student or the supervisor of the task. For parish activities, evidence usually consisted of a list of student signatures submitted by the supervisor.

Of the 38 who did not get confirmed, 25 did not have volunteering records on file. Since this rules out a large share of the unconfirmed students in the data, this paper cannot not use volunteering data as a predictor of program completion. It is unknown whether there was a rule for which student volunteering records were kept or not, so it is unknown whether the other 13 unconfirmed students with volunteering records were a random sample of all unconfirmed students.

Outside of the unconfirmed, there exists 11 more students who were confirmed, yet did not have a volunteering record. In lieu of volunteering records, a record of 0 volunteering events was assumed for these students.

*Student Absenteeism*

A student's absenteeism is calculated by adding 1 for each time a student was tardy marked, 1 for each time a student was marked leaving early, and 2 for each time a student was marked absent. Volunteer group instructors were in charge of making these marks. The marking of absences was rather consistent across group instructors, however the marking of tardiness and leaving early was less consistent. There is room for human error in this measurement.

In the case of severe weather or other unforeseen obstacles to attendance, classes were rescheduled. Rescheduled sessions may have higher absenteeism, which is a line of study for perhaps another paper.

The absence data provided only covered the second half of the program. Therefore, any students who dropped out of the program before the second half began would not have been considered in this study. To the knowledge of this researcher, only 6 such students exist.

## STATISTICAL ANALYSIS

### Estimation of True Mean

This study only looks at the students who attended the confirmation program between 2013 and 2016. If this is considered to be a random sample out of all the students who will ever participate in this program, then statistical inference can be used in order to estimate the true mean of this confirmation program. (Alternatively, this can be considered a random sample out of all the students who could have participated in this timespan. Interpretation would remain sound).

The dilemma of taking random samples is that it may not be representative of the population. So, there is the possibility that the sample mean is not close to the "actual" mean. Confidence intervals, a form of statistical inference, takes into account how the sample size and the desired confidence level to estimate a range where the population mean is likely to be.

For instance, consider the student volunteering for the 100 students who enrolled in the summer format from 2014 to 2016. On average, a summer student volunteered 5.74 times over the course of the program. If one wanted to go beyond 2014 to 2016 and estimate the average volunteering of all summer students, one would use a confidence interval.

A confidence interval is calculated with the formula $mean \pm z_{a/2} * \frac{s}{\sqrt{n}}$, where *mean* is the average value of the sample, *s* is the standard deviation of the population, *n* is the size of the sample, and *a* is the level of confidence desired. The standard deviation of the sample can be used in lieu of knowing the population standard deviation, however it can only work if $n > 30$. Note that the $z_{a/2}$ is the positive value which, when $\pm z_{a/2}$ are marked on the x-axis, bound a region under the standard normal curve with area $a$.

In the case of volunteering among the summer format students, the standard deviation is 2.74, the sample is size 100, and the mean is 5.74. Assuming a desired confidence of 95%, the normal curve reveals $z_{a/2} = 1.96$. So, the confidence interval is $5.74 \pm 1.96 * \frac{2.74}{\sqrt{100}}$, or $5.74 \pm 0.54$, or (5.2, 6.28). So, it can be concluded with 95% confidence that the true mean of student volunteering among summer format students is in the range (5.2, 6.28).

Sample means, standard deviations, and confidence intervals for each variable are provided for each cohort in Appendix A Table 2, 3, 4. By comparing these calculations to one another, some cohorts begin to stand out amongst the others. These differences will be studied in greater rigor later in the analysis, however the reader is encouraged to compare the Summer cohort with the program on average when looking at student absenteeism, the 2015 cohort with other years when looking at student volunteering, and the summer cohort again when looking at confirmation rates.

## Determining Correlations between Variables

When a researcher seeks to establish a relationship between two variables, he/she establishes a model for how changes in one variable influence changes in another. Then, the researcher compares the model to the actual data and determines how closely the model fits the data. A measure of this closeness is the correlation coefficient, commonly referred to as "*r*". This value varies from -1 to 1, with values closer to 0 indicating a poor fit and numbers further from 0 indicating a more accurate model. Another way of thinking about this is that an extreme *r*-value indicates a stronger correlation between the two variables through this model.

While formulas exist for finding the *r*-value between a model and the observed data, it is common practice to consult technology to determine the *r*-value due to the complexity of the formulas. After computing the *r*-value of a model with the data, the coefficient of determination is simply computed by squaring the *r*-value. This coefficient ranges from 0 to 1, with higher values indicating a stronger correlation between the variables. When interpreting a coefficient of determination, with the coefficient read as a percentage, one can claim that this share of the deviation in one variable can be accounted for by the changes in the other variable.

The most common model used by researchers is a linear model, because of how easy they are to reason with compared to most other models. In particular, the *r*-value not only indicates the strength of the relationship, but also the direction of the relationship. A positive *r*-value indicates a positive relationship, which means that for higher *x* values, one can expect higher *y* values. Likewise, a negative *r*-value indicates a negative relationship, which means that for higher *x* values, one can expect lower *y* values.

To illustrate the power of using a linear model, consider the variables of student volunteering and student absenteeism for each 568 students in this study. By fitting a line to the data, one can calculate how closely the model predicts the data. In this instance, the *r*-value of -0.27 is found. The coefficient of determination, 0.073, reveals that only 7.3% of the deviation in student volunteering can be attributed to deviations in absenteeism. This *r*-value is a weak correlation, so there is little reason to think a linear correlation exists between these variables.

While a strong linear correlation has not been found between these two variables, that does not mean a stronger correlation doesn't exist within certain cohorts. For instance, if this data was limited to include only 2014 students, the *r*-value becomes -0.34, which is stronger than the full program, yet still pretty weak. The *r*-values, and coefficients of determination, and interpretations between these student absenteeism and student volunteering for multiple cohorts are included in Table 1 (see Appendix B). On the whole, these relationships are quite weak. Thus, students with high volunteerism do not necessarily have low absenteeism.

After analyzing the relationship between student absenteeism and student volunteering, the relationship between student absenteeism and confirmation rates can be analyzed. Table 2 (see Appendix B) holds *r*-values, coefficients of determination, and interpretations for each cohort between student absenteeism and confirmation rates. On the whole, these relationships are even weaker than the relationships between student absenteeism and student volunteering.

Note that because of a bias in the data where most non-confirmed students did not have volunteering records on file, any analysis between these two variables will come out considerably stronger than what the true relationship may be. Therefore, these calculations are omitted from this study.

## Hypothesis Testing: Difference in Means

When estimating the true mean of each cohort, this paper urged the reader to notice the differences between the means of some of the cohorts. Sometimes, the differences between the means of these cohorts might be because of random chance, while sometimes the differences are true. Hypothesis testing is used to distinguish between these two situations.

In order to perform a hypothesis test, a researcher must first establish which hypothesis he/she wishes to challenge, which is called the null hypothesis. If the confidence in the null hypothesis being true is less than, say, 5%, then there is sufficient evidence that the null hypothesis is false and the opposite (the alternative hypothesis) is likely true. This is called rejecting the null hypothesis. However, if the confidence in the null hypothesis being true is greater than 5%, then there is insufficient evidence; the researcher has failed to reject the null hypothesis.

The 5% figure is called the level of significance. Values of 1%, 5%, and 10% are standard, with the smaller levels of significance indicating stronger evidence.

In this paper, the researcher seeks to challenge the statement that for a given variable and two cohorts, the true mean of one of the cohorts is less than or equal to the true mean of the other cohort. That is, given the sample mean of cohort A is lower than the sample mean of cohort B, the null hypothesis is $\mu_B \leq \mu_A$, where $\mu_A$ is the true mean of cohort A and $\mu_B$ is the true mean of cohort B. The researcher will reject the null hypothesis if there is less than 5% confidence in the null hypothesis.

To obtain how confident one is in this null hypothesis, the researcher must calculate the *t*-score, which is given by the formula $t = \frac{mean_A + \text{mean}_B}{\sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}}$, where $mean_A$ and $mean_B$ are the sample means of each dataset, $n_A$ and $n_B$ are the size of each dataset, and $s_A$ and $s_B$ are the standard deviations of each dataset. Use the true standard deviation if possible, otherwise use the sample standard deviation. Then, the researcher must calculate the degrees of freedom, which is approximated by the lesser of $n_A - 1$ and $n_B - 1$. Then, by consulting a *t*-distribution with the appropriate degrees of freedom, the researcher should look at *t* on the distribution, particularly the area to the right of it. The area given by this calculation is the confidence in the null hypothesis the researcher should have. This value is called the *p*-value.

Consider the Summer and Year cohorts when it comes to student volunteering. In this case, since the Summer cohort has a higher sample mean, the null hypothesis states that the mean of the Summer cohort is less than or equal to the mean of the Year cohort. The mean and standard deviation volunteerism for each cohort is provided in Appendix A, Table 3, while the count of each cohort is provided in Appendix A, Table 1. Then, the *t*-score can be calculated by following the formula: $t = \frac{5.74 - 5.32}{\sqrt{\frac{2.74^2}{100} + \frac{2.26^2}{468}}} = \frac{0.42}{\sqrt{0.075 + 0.011}} = \frac{0.42}{0.293} = 1.43$. Without rounding, this would be 1.42.

Before continuing, the degrees of freedom is 99, since *100 - 1 < 468 - 1*. Consulting technology to calculate the corresponding region under the *t* distribution with degrees of freedom being 99, the *p*-value (confidence in the null hypothesis) is 0.0787, or 7.87%. Since the 0.0787 was above the 0.05 (5%) significance level, this test has failed to reject the null hypothesis. So, there is insufficient evidence to show that the summer format has more volunteerism than the year-round format. However, it does look to be close, as it would pass at the 0.1 (10%) significance level.

The hypothesis tests which reject the null hypothesis at the 0.05 (5%) level are listed in Tables 1, 2, and 3 (see Appendix C), along with their corresponding confidences. Some of these results are discussed here.

The Summer cohort has a significantly lower confirmation rate, despite having lower absenteeism and slightly higher volunteerism. While this result is likely more vulnerable to random chance than the statistics have accounted for (see the Disclaimer section in the conclusion), this result does seem to indicate a commitment problem with summer students.

Another result is that 2015 and 2016 have lower than average confirmation rates, while 2014 has above average confirmation rates. Compared to the average student absenteeism, however, no major differences are observed. (2015 had lower absenteeism than 2013 and 2016, however not lower than the global average.)

Not Confirmed students had significantly higher absenteeism, as did students in the Low Vol. cohort. Likewise, Confirmed students and students in the High Vol. cohort had significantly lower absenteeism.

Meanwhile, 2015 had above average student volunteering. The statistics report Confirmed students having above average volunteering levels, however the average is brought down by the blank volunteering records assumed for the many unconfirmed students, so this result may not be trustworthy.

Students without volunteering records posted above average student absenteeism levels, and (not surprisingly) below average confirmation rates.

## CONCLUSION

The confirmation preparation program at St. Elizabeth Seton Church has a high confirmation rate (93%) for its young Catholics who complete the first half of the program.

The correlation between student absenteeism and confirmation rates ($r$ = -0.155) is of incredibly low strength, while the correlation between student absenteeism and student volunteering was also weak ($r$ = -0.272). Because of a bias in the collection of data, student volunteering and confirmation rates could not be compared.

Students who opted for the summer format exhibited significantly lower absenteeism, slightly higher student volunteering despite having less time and less reminders, yet significantly lower confirmation rates.

On a more general note, these data show there exists a remnant of students who go above and beyond to engage their faith at this age. Perhaps it is these teenagers who do not leave the church when they come of age; perhaps the future of the Catholic Church in the United States depends on these engaged teens.

## BIBLIOGRAPHY

Religion News article referenced in the introduction: http://religionnews.com/2015/09/16/catholic-us-catholics-measure/
Dynamic Catholic article referenced in the introduction: https://dynamiccatholic.com/sites/default/files/pdfs/Confirmation_Proposal_LR.pdf
Description of Catholic Confirmation program: http://stelizabethsetonchurch.org/youth-faith-formation

## APPENDICES

### A. Cohort Descriptions

**Table 1**: Cohort Descriptions. (An enumeration of the cohorts (subsets) analyzed in this study)

| Cohort Title | Criteria | Count |
|---|---|---|
| Full Sample | Attended class at least once | 568 |
| Summer | In Summer format | 100 |
| Year | In Year-round format | 468 |
| Confirmed | Confirmed year of class completion at St. Elizabeth Seton Church | 530 |
| Not Confirmed | Not Confirmed year of class completion or not at St. Elizabeth Seton Church | 38 |
| No Vol. | Did not have a volunteering record | 36 |
| Low Vol. | Number of volunteering events below 5-time guideline | 75 |
| High Vol. | Number of volunteering events above 8 | 39 |
| 2013 | Finished classes in 2013 | 151 |
| 2014 | Finished classes in 2014 | 146 |
| 2015 | Finished classes in 2015 | 135 |
| 2016 | Finished classes in 2016 | 136 |

**Table 2**: Student Absenteeism

| Cohort Title | Mean | Standard Deviation | Confidence Interval (95%) |
|---|---|---|---|
| Full Sample | 1.08 | 1.64 | [0.94, 1.21] |
| Summer | 0.06 | 0.34 | [0, 0.13] |
| Year | 1.29 | 1.73 | [1.14, 1.45] |
| Confirmed | 1.01 | 1.51 | [0.88, 1.14] |
| Not Confirmed | 2.03 | 2.78 | [1.14, 2.91] |
| No Vol. | 2.28 | 2.76 | [1.38, 3.18] |
| Low Vol. | 1.65 | 2.31 | [1.13, 2.18] |
| High Vol. | 0.28 | 0.79 | [0.03, 0.53] |
| 2013 | 1.26 | 1.39 | [1.04, 1.49] |
| 2014 | 0.99 | 1.49 | [0.75, 1.23] |
| 2015 | 0.81 | 1.46 | [0.57, 1.06] |
| 2016 | 1.22 | 2.14 | [0.86, 1.58] |

NOTE: This Table contains the data on mean, standard deviation, and 95% confidence intervals for each cohort on student absenteeism. Same information is provided for entire sample.

**Table 3**: Student Volunteering

| Cohort Title | Mean | Standard Deviation | Confidence Interval (95%) |
|---|---|---|---|
| Full Sample | 5.4 | 2.35 | [5.2, 5.59] |
| Summer | 5.74 | 2.74 | [5.2, 6.28] |
| Year | 5.32 | 2.26 | [5.12, 5.53] |
| Confirmed | 5.65 | 2.12 | [5.47, 5.83] |
| Not Confirmed | 1.84 | 2.65 | [1.0, 2.68] |
| 2013 | 5.19 | 2.17 | [4.85, 5.54] |
| 2014 | 5.28 | 1.63 | [5.02, 5.55] |
| 2015 | 5.93 | 2.18 | [5.56, 6.29] |
| 2016 | 5.22 | 3.18 | [4.69, 5.76] |

NOTE: This Table contains the data on mean, standard deviation, and 95% confidence intervals for each cohort on student volunteering. Left out “No Vol.”, “Low Vol.”, and “High Vol.” cohorts as results are skewed by definition. Same information is provided for the entire sample.

**Table 4**: Confirmation Rates

| Cohort Title | Mean | Standard Deviation | Confidence Interval (95%) |
|---|---|---|---|
| Full Sample | 93% | 0.25 | [91%, 95%] |
| Summer | 84% | 0.37 | [77%, 91%] |
| Year | 95% | 0.21 | [93%, 97%] |
| No Vol. | 31% | 0.47 | [15%, 46%] |
| Low Vol. | 99% | 0.12 | [96%, 100%] |
| High Vol. | 100% | 0.0 | [100%, 100%] |
| 2013 | 96% | 0.2 | [93%, 99%] |
| 2014 | 97% | 0.18 | [94%, 100%] |
| 2015 | 90% | 0.31 | [84%, 95%] |
| 2016 | 90% | 0.3 | [85%, 95%] |

NOTE: This Table contains the mean, standard deviation, and 95% confidence intervals for each cohort on confirmation rates. Left out the “Confirmed” and “Not Confirmed” cohorts, as results are obvious. Same information is provided for the entire sample.

## B. Correlations

**Table 1**: Student absenteeism and student volunteerism under a linear model

| Cohort Title | $r$ | $r^2$ | Interpretation |
|---|---|---|---|
| Full Sample | -0.27 | 0.074 | Weak: 7.4% of change accounted for. |
| Summer | -0.22 | 0.048 | Weak: 4.8% of change accounted for. |
| Year | -0.293 | 0.086 | Weak: 8.6% of change accounted for. |
| Confirmed | -0.221 | 0.049 | Weak: 4.9% of change accounted for. |
| Not Confirmed | -0.311 | 0.097 | Weak: 9.7% of change accounted for. |
| 2013 | -0.185 | 0.034 | Very Weak: 3.4% of change accounted for. |
| 2014 | -0.336 | 0.113 | Weak: 11.3% of change accounted for. |
| 2015 | -0.27 | 0.073 | Weak: 7.3% of change accounted for. |
| 2016 | -0.276 | 0.076 | Weak: 7.6% of change accounted for. |

NOTE: Displayed are $r$-values, coefficients of correlation, and interpretations for each cohort. The cohorts based on volunteering were excluded because they provided only a partial view of the volunteering data. Same information is provided for the entire sample.

**Table 2**: Student absenteeism and confirmation rates under a linear model

| Cohort Title | $r$ | $r^2$ | Interpretation |
|---|---|---|---|
| Full Sample | -0.155 | 0.024 | Very Weak: 2.4% of change accounted for. |
| Summer | -0.083 | 0.007 | Very Weak: 0.7% of change accounted for. |
| Year | -0.272 | 0.074 | Weak: 7.4% of change accounted for. |
| No Vol. | -0.2 | 0.04 | Weak: 4.0% of change accounted for. |
| Low Vol. | 0.084 | 0.007 | Very Weak: 0.7% of change accounted for. |
| High Vol. | - | - | 100% Confirmation rate - cannot calculate |
| 2013 | -0.132 | 0.017 | Very Weak: 1.7% of change accounted for. |
| 2014 | -0.282 | 0.079 | Weak: 7.9% of change accounted for. |
| 2015 | -0.077 | 0.006 | Very Weak: 0.6% of change accounted for. |
| 2016 | -0.189 | 0.036 | Very Weak: 3.6% of change accounted for. |

NOTE: Displayed are $r$-values, coefficients of correlation, and interpretations for each cohort. The cohorts based on confirmation were excluded because they provided only a partial view of the volunteering data. Same information is provided for the entire sample.

## C. Significantly Larger Means

**Table 1**: Confirmation Cohorts of Differing Means

| Lower Cohort Title | Lower Cohort Sample Mean | Greater Cohort Sample Mean | Greater Cohort Title | *p*-value (<0.05) |
|---|---|---|---|---|
| Summer | 0.84 | 0.933 | Full Sample | 0.008 |
| No Vol. | 0.306 | 0.933 | Full Sample | 0.0 |
| Full Sample | 0.933 | 0.987 | Low Vol. | 0.001 |
| Full Sample | 0.933 | 0.966 | 2014 | 0.039 |
| Summer | 0.84 | 0.953 | Year | 0.002 |
| No Vol. | 0.306 | 0.84 | Summer | 0.0 |
| Summer | 0.84 | 0.987 | Low Vol. | 0.0 |
| Summer | 0.84 | 0.96 | 2013 | 0.002 |
| Summer | 0.84 | 0.966 | 2014 | 0.001 |
| No Vol. | 0.306 | 0.953 | Year | 0.0 |
| Year | 0.953 | 0.987 | Low Vol. | 0.023 |
| 2015 | 0.896 | 0.953 | Year | 0.023 |
| 2016 | 0.904 | 0.953 | Year | 0.038 |
| No Vol. | 0.306 | 0.987 | Low Vol. | 0.0 |
| No Vol. | 0.306 | 0.96 | 2013 | 0.0 |
| No Vol. | 0.306 | 0.966 | 2014 | 0.0 |
| No Vol. | 0.306 | 0.896 | 2015 | 0.0 |
| No Vol. | 0.306 | 0.904 | 2016 | 0.0 |
| 2015 | 0.896 | 0.987 | Low Vol. | 0.002 |
| 2016 | 0.904 | 0.987 | Low Vol. | 0.003 |
| 2015 | 0.896 | 0.96 | 2013 | 0.02 |

| | | | | |
|---|---|---|---|---|
| 2016 | 0.904 | 0.96 | 2013 | 0.032 |
| 2015 | 0.896 | 0.966 | 2014 | 0.012 |
| 2016 | 0.904 | 0.966 | 2014 | 0.02 |

NOTE: An enumeration of the cohorts of greater means at the 0.05 (5%) significance level. The smaller and greater means for each cohort comparison are provided, along with the p-value rounded to the nearest thousandth. Since the Confirmed and Not Confirmed cohorts by definition have high and low means, they have been excluded from this table. Smaller *p*-values indicate stronger evidence.

**Table 2**: Absenteeism Cohorts of Differing Means

| Lower Cohort Title | Lower Cohort Sample Mean | Greater Cohort Sample Mean | Greater Cohort Title | *p*-value (<0.05) |
|---|---|---|---|---|
| Summer | 0.06 | 1.076 | Full Sample | 0.0 |
| Full Sample | 1.076 | 1.293 | Year | 0.02 |
| Full Sample | 1.076 | 2.026 | Not Confirmed | 0.022 |
| Full Sample | 1.076 | 2.278 | No Vol. | 0.007 |
| Full Sample | 1.076 | 1.653 | Low Vol. | 0.02 |
| High Vol. | 0.282 | 1.076 | Full Sample | 0.0 |
| 2015 | 0.815 | 1.076 | Full Sample | 0.035 |
| Summer | 0.06 | 1.293 | Year | 0.0 |
| Summer | 0.06 | 1.008 | Confirmed | 0.0 |
| Summer | 0.06 | 2.026 | Not Confirmed | 0.0 |
| Summer | 0.06 | 2.278 | No Vol. | 0.0 |
| Summer | 0.06 | 1.653 | Low Vol. | 0.0 |
| Summer | 0.06 | 0.282 | High Vol. | 0.05 |
| Summer | 0.06 | 1.265 | 2013 | 0.0 |
| Summer | 0.06 | 0.986 | 2014 | 0.0 |
| Summer | 0.06 | 0.815 | 2015 | 0.0 |
| Summer | 0.06 | 1.221 | 2016 | 0.0 |
| Confirmed | 1.008 | 1.293 | Year | 0.003 |
| Year | 1.293 | 2.278 | No Vol. | 0.021 |
| High Vol. | 0.282 | 1.293 | Year | 0.0 |
| 2014 | 0.986 | 1.293 | Year | 0.019 |
| 2015 | 0.815 | 1.293 | Year | 0.001 |

| | | | | |
|---|---|---|---|---|
| Confirmed | 1.008 | 2.026 | Not Confirmed | 0.016 |
| Confirmed | 1.008 | 2.278 | No Vol. | 0.005 |
| Confirmed | 1.008 | 1.653 | Low Vol. | 0.011 |
| High Vol. | 0.282 | 1.008 | Confirmed | 0.0 |
| Confirmed | 1.008 | 1.265 | 2013 | 0.026 |
| High Vol. | 0.282 | 2.026 | Not Confirmed | 0.0 |
| 2014 | 0.986 | 2.026 | Not Confirmed | 0.016 |
| 2015 | 0.815 | 2.026 | Not Confirmed | 0.007 |
| High Vol. | 0.282 | 2.278 | No Vol. | 0.0 |
| 2013 | 1.265 | 2.278 | No Vol. | 0.02 |
| 2014 | 0.986 | 2.278 | No Vol. | 0.005 |
| 2015 | 0.815 | 2.278 | No Vol. | 0.002 |
| 2016 | 1.221 | 2.278 | No Vol. | 0.019 |
| High Vol. | 0.282 | 1.653 | Low Vol. | 0.0 |
| 2014 | 0.986 | 1.653 | Low Vol. | 0.013 |
| 2015 | 0.815 | 1.653 | Low Vol. | 0.003 |
| High Vol. | 0.282 | 1.265 | 2013 | 0.0 |
| High Vol. | 0.282 | 0.986 | 2014 | 0.0 |
| High Vol. | 0.282 | 0.815 | 2015 | 0.002 |
| High Vol. | 0.282 | 1.221 | 2016 | 0.0 |
| 2014 | 0.986 | 1.265 | 2013 | 0.049 |
| 2015 | 0.815 | 1.265 | 2013 | 0.004 |
| 2015 | 0.815 | 1.221 | 2016 | 0.035 |

NOTE: An enumeration of the cohorts of greater means at the 0.05 (5%) significance level. The smaller and greater means for each cohort comparison are provided, along with the p-value rounded to the nearest thousandth. Smaller p-values indicate stronger evidence.

**Table 3**: Volunteering Cohorts of Differing Means

| **Lower Cohort Title** | **Lower Cohort Sample Mean** | **Greater Cohort Sample Mean** | **Greater Cohort Title** | ***p*-value (<0.05)** |
|---|---|---|---|---|
| Full Sample | 5.396 | 5.651 | Confirmed | 0.03 |
| Full Sample | 5.396 | 5.926 | 2015 | 0.007 |
| Not Confirmed | 1.842 | 5.74 | Summer | 0.0 |
| 2013 | 5.192 | 5.74 | Summer | 0.048 |
| Year | 5.323 | 5.651 | Confirmed | 0.009 |
| Year | 5.323 | 5.926 | 2015 | 0.003 |
| 2013 | 5.192 | 5.651 | Confirmed | 0.011 |
| 2014 | 5.281 | 5.651 | Confirmed | 0.013 |
| 2013 | 5.192 | 5.926 | 2015 | 0.003 |
| 2014 | 5.281 | 5.926 | 2015 | 0.003 |
| 2016 | 5.221 | 5.926 | 2015 | 0.017 |

NOTE: An enumeration of the cohorts of greater means at the 0.05 (5%) significance level. The smaller and greater means for each cohort comparison are provided, along with the *p*-value rounded to the nearest thousandth. Volunteering-based cohorts (Low, High, and No Vol. cohorts) have been excluded, as they are naturally higher/lower than other cohorts. Also, since many of the unconfirmed students were assumed to have no volunteering record, unconfirmed students are expected to have lower volunteering means anyway. Smaller *p*-values indicate stronger evidence.

## D. Source Code

The researcher of this paper developed code to be run to assist in the managing of data, crunching of numbers, and writing of this paper. This code is available on GitHub. The author makes no assurances that this code will work on other computers, as some shortcuts in terms of file paths were made.

The software used to handle datasets and conduct analysis is a Python package being developed by the researcher called analytics. The version of this software is included with the software specific to this project was also developed with Python. The code is available at https://github.com/jamesfulford/math-capstone.

* **JAMES PATRICK FULFORD** is a bright young problem solver who credits God for his achievements. He is interested in solving problems using statistics and machine learning, paired with critical thinking the power of computers. His hobbies include strategy games, programming in Python, and being out in the woods. After he graduates with his B.A. in Mathematics from Rivier University, he plans on attending the University of New Hampshire for his Masters in Computer Science.