Analysis of Souls

# A Statistical Study of Confirmation Students

## By James Fulford Fall 2016 - Spring 2017

When an educational program is run, there are (at least) three key metrics to track success: attendance, engagement, and completion. In many cases, these metrics are linked: lower attendance tends to predict lower achievement and lower graduation rates. However, this relationship is not always the case.  
  
 Take, for instance, the Catholic Confirmation program at St. Elizabeth Seton Parish in Bedford, New Hampshire. From 2013 to 2016, a total of 568 students high school sophomores attended the confirmation program. The purpose of this program is to educate students about their Catholic faith and prepare them for the sacrament of Confirmation. In this program, graduation is choosing to join the Catholic Church by receiving the sacrament of Confirmation. To prepare, these students are required to attend classes and are pushed to engage the community by volunteering at least five times for the parish. These three metrics (confirmation, attendance, and graduation) are analyzed in this paper.

## Confirmation

A student is considered confirmed if they attended the St. Elizabeth Seton confirmation ceremony the same year the final class occurred. The reason for this strict definition is because some students delayed by a year or were confirmed in another parish. In general, the program had unusually high completion rates.  
  
 In total, 530 students were confirmed (93.3%) and 38 were not confirmed (6.7%).

## Absenteeism

The seven classes were offered in two formats: summer and year-round. The summer option was added in 2014, and a total of 100 students went for this option (17.6%). The year-round format hosted a total of 468 students (82.4%).  
  
 Year-round students met on Sunday afternoons once a month, seven times a school year, for two school years. The scope of this study only includes students who at least started the second year of the program. No such half-way cut-off exists for summer students.  
  
 The method for counting absences is one point for every day a teacher marks a student as tardy or as an early-leaver, and 2 points for full absences.

## Volunteering

What qualifies as a volunteering event is left to the discretion of the program coordinator. Mostly, volunteering events consisted of visiting nursing homes, committing hours at food pantries/soup kitchens, or helping with parish events (dinners, youth programs, etc.) Longer volunteering events occasionally counted for more.  
  
 Of the 38 who did not get confirmed, 25 did not have volunteering records on file. This hole in the data is why this paper cannot not use volunteering data as a predictor of program completion. The 13 that did have volunteering records may have been confirmed in later years or in different parishes.

## Cohorts

When doing this analysis, 12 subsets of the dataset were also analyzed to see if they behaved differently compared to one another and the full dataset. These cohorts are enumerated in the following table:

|  |  |  |
| --- | --- | --- |
| Cohort Title | Criteria | Count |
| Base | Attended program within year of Confirmation ceremony | 568 |
| Summer | In Summer program | 100 |
| Year | In Year-round program | 468 |
| Confirmed | Confirmed year of class completion at St. Elizabeth Seton | 530 |
| Not Confirmed | Not Confirmed year of class completion or not at St. Elizabeth Seton | 38 |
| No Vol. | Did not have a volunteering record | 36 |
| Low Vol. | Number of volunteering events below 5-event 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 |

# Centers of Distributions

## Confirmation

The mean is 0.933. At a 0.95 level of confidence, the population mean is within the interval of [0.91, 0.95].

### Confirmation: Summer

The mean is 0.84. At a 0.95 level of confidence, the population mean is within the interval of [0.77, 0.91].

### Confirmation: Confirmed

The mean is 1.0. At a 0.95 level of confidence, the population mean is within the interval of [1.0, 1.0].

### Confirmation: Not Confirmed

The mean is 0.0. At a 0.95 level of confidence, the population mean is within the interval of [0.0, 0.0].

### Confirmation: No Vol.

The mean is 0.306. At a 0.95 level of confidence, the population mean is within the interval of [0.15, 0.46].

### Confirmation: Low Vol.

The mean is 0.987. At a 0.95 level of confidence, the population mean is within the interval of [0.96, 1.01].

### Confirmation: High Vol.

The mean is 1.0. At a 0.95 level of confidence, the population mean is within the interval of [1.0, 1.0].

### Confirmation: 2014

The mean is 0.966. At a 0.95 level of confidence, the population mean is within the interval of [0.94, 1.0].

## Absenteeism

The mean is 1.076. At a 0.95 level of confidence, the population mean is within the interval of [0.94, 1.21].

### Absenteeism: Summer

The mean is 0.06. At a 0.95 level of confidence, the population mean is within the interval of [-0.01, 0.13].

### Absenteeism: Year

The mean is 1.293. At a 0.95 level of confidence, the population mean is within the interval of [1.14, 1.45].

### Absenteeism: Not Confirmed

The mean is 2.026. At a 0.95 level of confidence, the population mean is within the interval of [1.14, 2.91].

### Absenteeism: No Vol.

The mean is 2.278. At a 0.95 level of confidence, the population mean is within the interval of [1.38, 3.18].

### Absenteeism: Low Vol.

The mean is 1.653. At a 0.95 level of confidence, the population mean is within the interval of [1.13, 2.18].

### Absenteeism: High Vol.

The mean is 0.282. At a 0.95 level of confidence, the population mean is within the interval of [0.03, 0.53].

### Absenteeism: 2015

The mean is 0.815. At a 0.95 level of confidence, the population mean is within the interval of [0.57, 1.06].

## Volunteering

The mean is 5.396. At a 0.95 level of confidence, the population mean is within the interval of [5.2, 5.59].

### Volunteering: Confirmed

The mean is 5.651. At a 0.95 level of confidence, the population mean is within the interval of [5.47, 5.83].

### Volunteering: Not Confirmed

The mean is 1.842. At a 0.95 level of confidence, the population mean is within the interval of [1.0, 2.68].

### Volunteering: No Vol.

The mean is 0.0. At a 0.95 level of confidence, the population mean is within the interval of [0.0, 0.0].

### Volunteering: Low Vol.

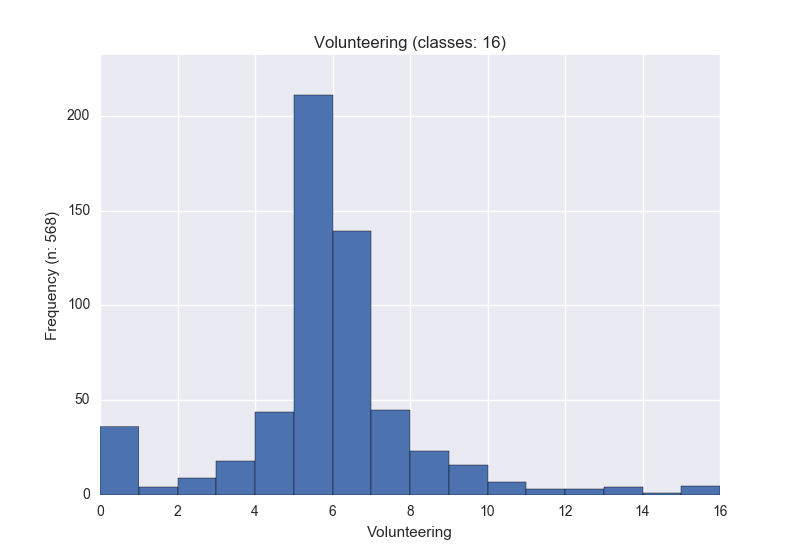
The mean is 3.36. At a 0.95 level of confidence, the population mean is within the interval of [3.16, 3.56].

### Volunteering: High Vol.

The mean is 10.897. At a 0.95 level of confidence, the population mean is within the interval of [10.2, 11.59].

### Volunteering: 2015

The mean is 5.926. At a 0.95 level of confidence, the population mean is within the interval of [5.56, 6.29].



# Variable Relationships

## Confirmation and Volunteering: Base

The coefficient of linear correlation is 0.405.  
  
 Approximately 16.4% of the variation in Confirmation can be accounted for by variations in Volunteering. Similarly, 16.4% of the variation in Volunteering can be accounted for by variations in Confirmation.

## Confirmation and Volunteering: Summer

The coefficient of linear correlation is 0.439.  
  
 Approximately 19.3% of the variation in Confirmation can be accounted for by variations in Volunteering. Similarly, 19.3% of the variation in Volunteering can be accounted for by variations in Confirmation.

## Confirmation and Volunteering: Year

The coefficient of linear correlation is 0.425.  
  
 Approximately 18.1% of the variation in Confirmation can be accounted for by variations in Volunteering. Similarly, 18.1% of the variation in Volunteering can be accounted for by variations in Confirmation.

## Confirmation and Volunteering: 2013

The coefficient of linear correlation is 0.411.  
  
 Approximately 16.9% of the variation in Confirmation can be accounted for by variations in Volunteering. Similarly, 16.9% of the variation in Volunteering can be accounted for by variations in Confirmation.

## Confirmation and Volunteering: 2014

The coefficient of linear correlation is 0.495.  
  
 Approximately 24.5% of the variation in Confirmation can be accounted for by variations in Volunteering. Similarly, 24.5% of the variation in Volunteering can be accounted for by variations in Confirmation.

## Confirmation and Volunteering: 2015

The coefficient of linear correlation is 0.324.  
  
 Approximately 10.5% of the variation in Confirmation can be accounted for by variations in Volunteering. Similarly, 10.5% of the variation in Volunteering can be accounted for by variations in Confirmation.

## Confirmation and Volunteering: 2016

The coefficient of linear correlation is 0.488.  
  
 Approximately 23.8% of the variation in Confirmation can be accounted for by variations in Volunteering. Similarly, 23.8% of the variation in Volunteering can be accounted for by variations in Confirmation.

## Absenteeism and Volunteering: Not Confirmed

The coefficient of linear correlation is -0.311.  
  
 Approximately 9.7% of the variation in Absenteeism can be accounted for by variations in Volunteering. Similarly, 9.7% of the variation in Volunteering can be accounted for by variations in Absenteeism.

## Absenteeism and Volunteering: 2014

The coefficient of linear correlation is -0.336.  
  
 Approximately 11.3% of the variation in Absenteeism can be accounted for by variations in Volunteering. Similarly, 11.3% of the variation in Volunteering can be accounted for by variations in Absenteeism.

## Confirmation

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Lower Cohort | Lower Mean | Greater Mean | Greater Cohort | p value |
| Summer | 0.84 | 0.933 | Base | 0.008 |
| No Vol. | 0.306 | 0.933 | Base | 0.0 |
| Base | 0.933 | 0.987 | Low Vol. | 0.001 |
| Base | 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 |

## Absenteeism

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Lower Cohort | Lower Mean | Greater Mean | Greater Cohort | p value |
| Summer | 0.06 | 1.076 | Base | 0.0 |
| Base | 1.076 | 1.293 | Year | 0.02 |
| Base | 1.076 | 2.026 | Not Confirmed | 0.022 |
| Base | 1.076 | 2.278 | No Vol. | 0.007 |
| Base | 1.076 | 1.653 | Low Vol. | 0.02 |
| High Vol. | 0.282 | 1.076 | Base | 0.0 |
| 2015 | 0.815 | 1.076 | Base | 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 |

## Volunteering

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Lower Cohort | Lower Mean | Greater Mean | Greater Cohort | p value |
| Base | 5.396 | 5.651 | Confirmed | 0.03 |
| Not Confirmed | 1.842 | 5.396 | Base | 0.0 |
| Low Vol. | 3.36 | 5.396 | Base | 0.0 |
| Base | 5.396 | 10.897 | High Vol. | 0.0 |
| Base | 5.396 | 5.926 | 2015 | 0.007 |
| Not Confirmed | 1.842 | 5.74 | Summer | 0.0 |
| Low Vol. | 3.36 | 5.74 | Summer | 0.0 |
| Summer | 5.74 | 10.897 | High Vol. | 0.0 |
| 2013 | 5.192 | 5.74 | Summer | 0.048 |
| Year | 5.323 | 5.651 | Confirmed | 0.009 |
| Not Confirmed | 1.842 | 5.323 | Year | 0.0 |
| Low Vol. | 3.36 | 5.323 | Year | 0.0 |
| Year | 5.323 | 10.897 | High Vol. | 0.0 |
| Year | 5.323 | 5.926 | 2015 | 0.003 |
| Not Confirmed | 1.842 | 5.651 | Confirmed | 0.0 |
| Low Vol. | 3.36 | 5.651 | Confirmed | 0.0 |
| Confirmed | 5.651 | 10.897 | High Vol. | 0.0 |
| 2013 | 5.192 | 5.651 | Confirmed | 0.011 |
| 2014 | 5.281 | 5.651 | Confirmed | 0.013 |
| Not Confirmed | 1.842 | 3.36 | Low Vol. | 0.001 |
| Not Confirmed | 1.842 | 10.897 | High Vol. | 0.0 |
| Not Confirmed | 1.842 | 5.192 | 2013 | 0.0 |
| Not Confirmed | 1.842 | 5.281 | 2014 | 0.0 |
| Not Confirmed | 1.842 | 5.926 | 2015 | 0.0 |
| Not Confirmed | 1.842 | 5.221 | 2016 | 0.0 |
| Low Vol. | 3.36 | 10.897 | High Vol. | 0.0 |
| Low Vol. | 3.36 | 5.192 | 2013 | 0.0 |
| Low Vol. | 3.36 | 5.281 | 2014 | 0.0 |
| Low Vol. | 3.36 | 5.926 | 2015 | 0.0 |
| Low Vol. | 3.36 | 5.221 | 2016 | 0.0 |
| 2013 | 5.192 | 10.897 | High Vol. | 0.0 |
| 2014 | 5.281 | 10.897 | High Vol. | 0.0 |
| 2015 | 5.926 | 10.897 | High Vol. | 0.0 |
| 2016 | 5.221 | 10.897 | High Vol. | 0.0 |
| 2013 | 5.192 | 5.926 | 2015 | 0.003 |
| 2014 | 5.281 | 5.926 | 2015 | 0.003 |
| 2016 | 5.221 | 5.926 | 2015 | 0.017 |

## References

@Article{Hunter:2007,  
 Author = {Hunter, J. D.},  
 Title = {Matplotlib: A 2D graphics environment},  
 Journal = {Computing In Science & Engineering},  
 Volume = {9},  
 Number = {3},  
 Pages = {90--95},  
 abstract = {Matplotlib is a 2D graphics package used for Python  
 for application development, interactive scripting, and  
 publication-quality image generation across user  
 interfaces and operating systems.},  
 publisher = {IEEE COMPUTER SOC},  
 doi = {10.1109/MCSE.2007.55},  
 year = 2007  
}