Final project

Swetha Chamala

Bellevue University

**Final project**

**Introduction**

Top of Form

With a prospective kindergartener in my household my mission since last couple of months is to analyze the school system in New York city. Naturally, I began to gather as much data and information as I can to learn about public, private and charter schools in the New York city (NYC) area. During my research one such aspect that piqued my curiosity is 'School safety'. Being the cultural, financial and media capital of the world, how safe are the public schools in NYC?

It is reported that the New York City Police Department (NYPD) has been tasked with the collection and maintenance of crime data for incidents that occur in New York City Public schools. The NYPD has provided this data to the New York City Department of Education (DOE) and DOE has compiled this data by schools and locations for the information of the general public. Listed below are the datasets that I chose to work with -

Dataset 1: [NY 2010 - 2016 School safety report](https://data.cityofnewyork.us/Education/2010-2016-School-Safety-Report/qybk-bjjc)

This data set has 34 columns and 6310 rows

Dataset 2: [2016 - 2017 School safety report](https://data.cityofnewyork.us/Education/2016-2017-School-Safety-Report/rear-wh5i)

This data set has 33 columns and 2046 rows

**Hypothesis**

Several news articles suggest that there is increase of crimes in the city's public schools. Especially in the school year 2017. So, I will be testing the hypothesis that the school year 2016-17 has higher crime rates reported than previous years. The null hypothesis would state the opposite. In addition to testing the hypothesis, my aim for this study is to identify trends in the total crimes with each school year. Whether there is an increase or decrease in the crime rate. If there is a trend, what sort of crimes occur the most - violent crimes, non-criminal cases, property crimes or other crimes. Find any anomalies if any.

**Variables**

I chose Major N, Oth N, NoCrim N, Prop N, Vio N as variables.

* School year – is the academic year of school
* Borough – New York City borough the school is situated in
* # Schools – Number of schools located in the same building
* Geographical district code – The school’s geographical district as defined by the NYC Department of Education
* Register – Number of students on register
* Major N - Number of major crimes
* Oth N – Number of ‘other’ crimes
* NoCrim N – Number of non-criminal crimes
* Prop N – Number of property related crimes
* Vio N – Number of violent crimes
* RangeA – building population
* AvgOfMajor N - average of major crimes for all buildings that have the same EnGroupA/Range A
* AvgOfOth N - average of other crimes for all buildings that have the same EnGroupA/Range A
* AvgOfNoCrim N - average of non-criminal crimes for all buildings that have the same EnGroupA/Range A
* AvgOfProp N - average of property crimes for all buildings that have the same EnGroupA/Range A
* AvgOfVio N - average of violent crimes for all buildings that have the same EnGroupA/Range A

**Histograms**

Dataset 1:



The 5 variables - Major\_N, NoCrim\_N, Oth\_N, Prop\_N, Vio\_N have similar distributions with smaller values. Thereby same trend is observed in their corresponding 'Avg' distributions.

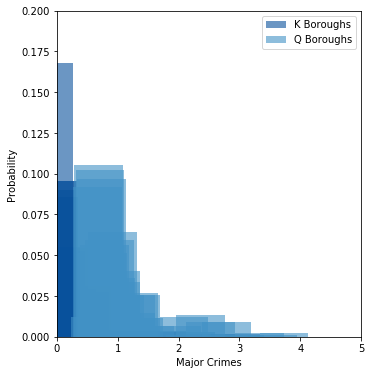
Dataset 2:



Similar to dataset 1, the 2016-2017 dataset 2 also follows the same trends. The 5 variables - Major\_N, NoCrim\_N, Oth\_N, Prop\_N, Vio\_N have similar distributions. Thereby same trend is observed in their corresponding 'Avg' distributions.

**PMFs**

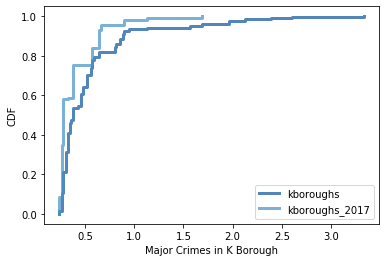
Plotting PMFs of ‘AvgOfMajor\_N’ grouped by K (Brooklyn) and Q (Queens) boroughs as an example of PMF plots shown in page 29 of Thinkstats book



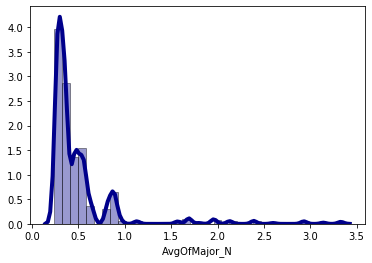
**CDFs**

Plotting CDFs of ‘Major\_N’ column of dataset 1 (2010-2016) and dataset 2 (2016-2017)

The plot shows that the reported major crimes are higher for 2010-16 compared to 2016-17

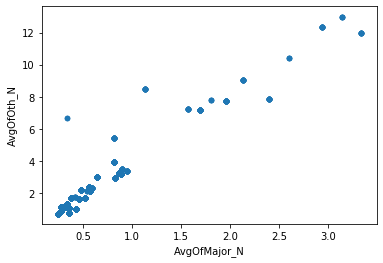


**Analytical distribution**

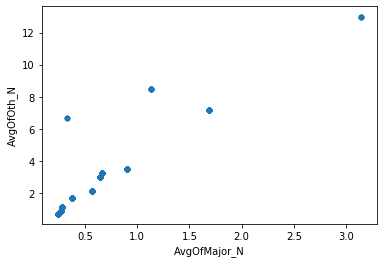
****

Plot of ‘AvgofMajor\_N’ of dataset 1

**Scatter plots**

****

Comparing tow variables AvgOfMajor\_N and AvgOfOth\_N of dataset 1



Comparing tow variables AvgOfMajor\_N and AvgOfOth\_N of dataset 2

Both the scatter plots show that there is a positive linear relationship between two variables. However, datsset 1 suggests there is strong correlation compared to dataset 2

**Testing hypothesis**

I tested hypothesis by calculating means and differences.

Means of ‘all types of crimes’ of dataset 1 grouped by school year

|  |  |  |  |
| --- | --- | --- | --- |
| 2013-14 | 0.0 | 126.0 | 8.102979 |
| 2014-15 | 0.0 | 128.0 | 7.950169 |
| 2015-16 | 0.0 | 130.0 | 7.711149 |

Mean of ‘all types of crimes’ of dataset 2 grouped by school year

|  |  |  |  |
| --- | --- | --- | --- |
| 2016 - 17 | 0.0 | 114.0 | 7.485501 |

Performing ttest:

Ttest\_indResult(statistic=1.0085891763517827, pvalue=0.31322309609765847)

Total crime rates of the year 2017 are not higher compared to previous years.

P-value > 0.05 suggests null-hypothesis is true. Which proves my hypothesis wrong. Implying that school crimes are not on the rise compared to previous years.

References:

https://www.nydailynews.com/new-york/education/crime-nyc-public-schools-nypd-data-shows-article-1.3851995

https://www.nyclu.org/en/press-releases/nypd-releases-complete-school-safety-data-first-time

https://www.the74million.org/article/exclusive-how-safe-are-nycs-schools-new-interactive-map-compares-what-teachers-students-are-reporting/

https://www.wnyc.org/story/school-safety-incidents-vary-depending-who-counts/

Bottom of Form