Analysis of Houses and Apartments in relation to Toronto B & E Crimes

University of Toronto

Yuan Ji, Roger Lam, Yi Yang, Tianqi Sun, TUT0205-1

March 30, 2020

Introduction

Break and enter robberies are one of the most common crimes in Toronto, endangering and disturbing the lives of many civilians. We believe that the blueprint to handling B & E crimes is closely related to the premise type of where these crimes occur.

Therefore our research question is: **How do houses and apartments differ among B & E crimes?**

Prerequisites

There are two fundamental subquestions needed to answer this:

1. What is the relationship between the offences to houses and apartments and the overall B & E rates?
2. How does the mean difference of reported and occurrence time of B & E robberies relate between houses and apartments?

What is the relationship between the offences to houses and apartments and the overall B & E rates?

Corresponding Datasets

{We used the TPS B & E dataset, a subset of the Major Crime Indicators (MCI) 2014 to 2019 occurrences (by reported date), and their Neighbourhood Crime Rates dataset, a 2014 to 2019 boundary file of crime data per neighbourhood.} {From these two datasets, we created and used:}

- {area_dataset - dataset of which divided neighbourhoods into small groups based on their neighbourhood IDs}
- {crime_rate_dataset - dataset of which outlines B & E robberies per neighbourhood group}

{Note that only incidences of which the crime was reported at most a year after the occurrence were observed for sensible results. Further, we have filtered our data to represent observations pertinent to houses and apartments exclusively.}

Important Variables

{These are the variables that were crucial to answering our question:}

- {area_num (created categorical variable): group ID number assigned to neighbourhoods within intervals of 10 of their inherent IDs.}
- {time_difference (created numerical variable): the time difference in hours between the occurrence date of the B & E crime and its' reported date.}
- {num_b_e_crimes (created numerical variable): the number of B & E crimes of the corresponding neighbourhood group based on ID area_num.}
- {premisetype (given categorical variable: indication of whether the crime occurred in a house, apartment, commercial building, or other.}
- {mean_crime_rate (created numerical variable): the mean crime rate of the corresponding neighbourhood group based on ID area_num.}

Statistical Methods

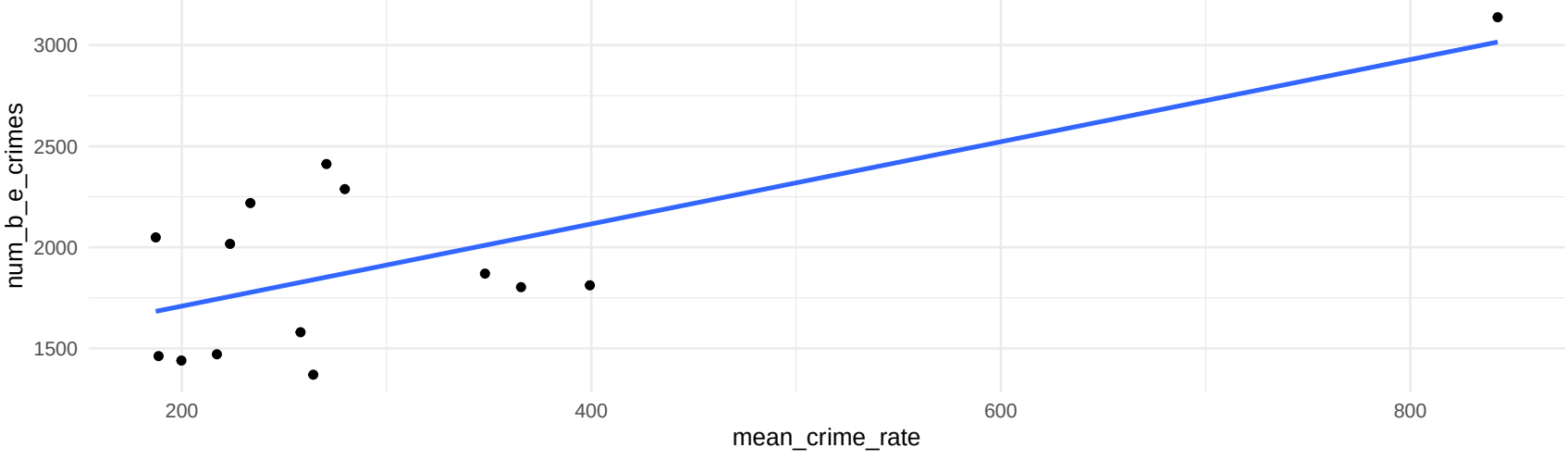
Our approach was a statistical method called linear regression, a data analysis technique focused on finding the relationship between a response variable and an explanatory variable.

With this method, we were able to investigate the relationship between our created variables num_b_e_crimes and mean_crime_rate.

Firstly, we plotted number of B & E crimes in relation to the mean crime rate to a scatterplot. Then, used the variable mean_crime_rate as an explanatory variable to the variable num_b_e_crimes to acquire coefficients that provide us a fitted linear regression model; a linear representation of our data points.

Results

For the scatter plot we produced, all points except one fall between mean_crime_rate equals 200 and 400. Although these points do not show a very tight relation between the number of crimes and mean crime rates, we can still that these two variables are positively related. One outstanding outlier has almost double the averge number of B & E robberies as well as triple the average mean crime rate, this is the data point of group ID number 8.



Results cont'd

As for the linear regression model, our calculated coefficients produce the equation:

$$\text{num\_b\_e\_crimes} = 1303 + 2.033 * \text{mean\_crime\_rate}$$

This equation predicts that for each unit change in the mean crime rate implies that two extra B & E crimes occur.

The calculated correlation (indication of strength of the relationship) is around 0.70, which means there is a moderately strong relationship. On the other hand, the calculated coefficient of determination is around 0.50, which tells us that only half of the values of num_b_e_crimes can be explained from our model.

Implications

A correlation of 0.70 and a coefficient of determination of 0.50 indicates that there is somewhat of a relationship between mean crime rates and total number of crimes taken place in both apartments and houses.

Higher crime rates, may or may not indicate larger numbers of B & E robberies.

Living in houses and living in apartments are both under the risk of B & E robberies, the same precautions need to be taken for people living in either apartments or houses.

Implications cont'd

{Nonetheless, there is one key takeaway shown by outlier, data point of group ID number 8. This group of neighbourhoods includes:}

Bay Sreet Corridor, Church-Yonge Corridor, North St. James Town, Waterfront Communities - The Island, University, Kensington-Chinatown, Palmerston, Little Italy, Regent Park, Moss Park, Cabbagetown - South St. James Town

{These are the neighbourhoods that accumulate to an abnormally high number of B & E crimes as well as the highest crime rates in general. By our statistics, we reccommend the law monitor these neighbourhoods more often.}

How does the mean difference of reported and occurrence time of B & E robberies relate between houses and apartments?

Corresponding Datasets

{We used the TPS B & E dataset, a subset of the Major Crime Indicators (MCI) 2014 to 2019 occurrences (by reported date). From this dataset, we created and used:}

- {time_dataset - dataset of which summarised the time related details of the B & E dataset}
- {house_dataset - dataset of which contains B & E data pertinent only to houses}
- {apt_dataset - dataset of which contains B & E data pertinent only to apartments}

{Note that only incidences of which the crime was reported at most a year after the occurrence were observed for sensible results. Further, we have filtered our data to represent observations pertinent to houses and apartments exclusively.}

Important Variables

{These are the variables that were crucial to answering our question:}

- {area_num (created categorical variable): group ID number assigned to neighbourhoods within intervals of 10 of their inherent IDs.}
- {reporteddate (given numerical variable): date of crime occurrence}
- {occurrencedate (given numerical variable): date crime occurrence was reported}
- {time_difference (created numerical variable): the time difference in hours between the occurrence date of the B & E crime and its' reported date.}
- {premisetype (given categorical variable: indication of whether the crime occurred in a house, apartment, commercial building, or other.}
- {mean_td (created numerical variable): mean time difference in hours of the corresponding neighbourhood group based on ID area_num}

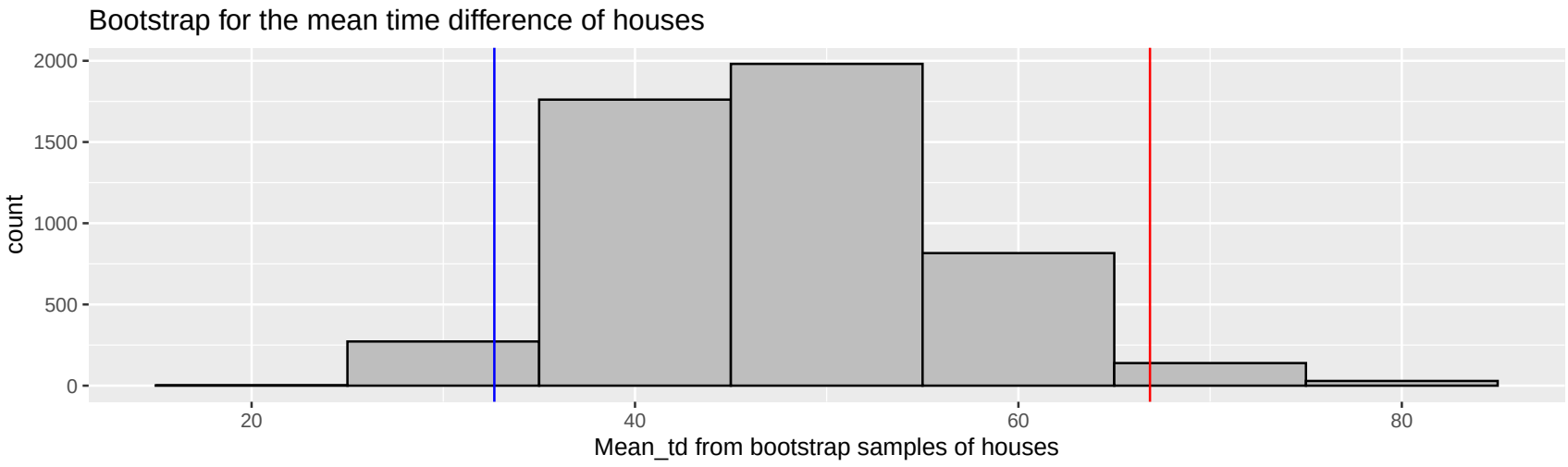
Statistical Methods

We used a statistical method called bootstrapping, a parameter analysis technique that uses random sampling of a representative statistic to find results.

The fact that the TPS's dataset is representative of the Toronto population, we used the mean of our variable time_difference as our original test statistic. We then resampled by maintaining observational values and swapping labels, under a simulation of 5000 repetitions.

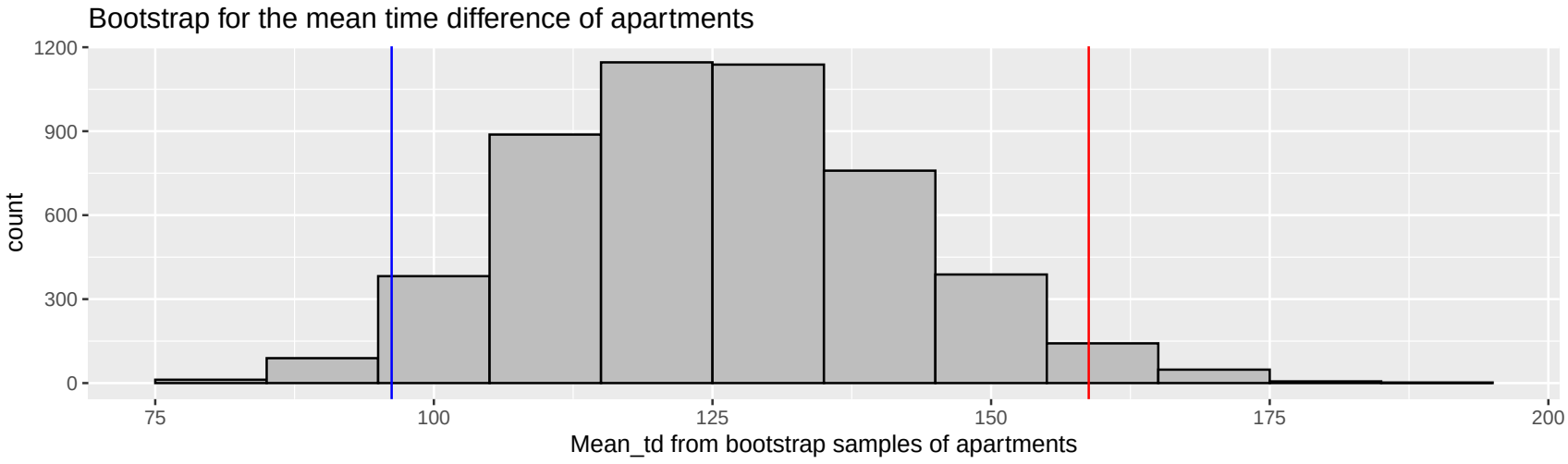
This creates a sampling distribution on a histogram, where we can find an interval of which we are 95% confident the actual mean time difference is in (known as a confidence interval).

mean_td Sampling Distribution for Houses




```
##      2.5%     97.5%
## 32.66204 66.86312
```

mean_td Sampling Distribution for Apartments




```
##       2.5%      97.5%
##   96.21148 158.75975
```

Results

Confidence interval for houses: (32.66204, 66.86312)

Confidence interval for apartments: (96.21148, 158.75975)

Therefore...

- We are 95% confident that the mean time difference for all houses residents of reporting B & E offence and occurrence of B & E offence is approximatley between 32.66 hours and 66.86 hours.
- We are 95% confident that the mean time difference for all apartment residents of reporting B & E offence and occurrence of B & E offence is approximatley between 96.21 hours and 158.76 hours.

Results cont'd

We also find out that there's a large difference between these two intervals, which indicates that residents in apartments are usually less responsive to B & E crimes compared to those living in houses. Moreover, the response time of apartment residents is more spread out than those living in houses.

Implications

Since apartment residents are usually less responsive to B & E crimes, we reccommend that the TPS enforce regulation around neighbourhoods of apartments, improve safety, and antitheft systems of each apartment. In addition, seek out ways to educate apartment dwellers on how to prevent and handle these situations.

From these results, we conclude that apartment residents take longer times to respond to and report B & E offences than those living in houses.

How do houses and apartments differ among B & E crimes?

Corresponding Datasets

{We used the TPS B & E dataset, a subset of the Major Crime Indicators (MCI) 2014 to 2019 occurrences (by reported date), and their Neighbourhood Crime Rates dataset, a 2014 to 2019 boundary file of crime data per neighbourhood.} {From these two datasets, we created and used:}

- {area_dataset - dataset of which divided neighbourhoods into small groups based on their neighbourhood IDs}
- {crime_rate_dataset - dataset of which outlines B & E robberies per neighbourhood group}

{For each of our grouped areas, we calculated the numbers of B & E of houses and apartments; then, we calculated the population, total area and population density for each grouped ID number.}

Important Variables

{These are the variables that were crucial to answering our question:}

- {area_num (created numerical variable): group ID number assigned to neighbourhoods within intervals of 10 of their inherent IDs.}
- {mean_b_and_e_rate (created numerical variable): indication of whether the crime occurred in a house, apartment, commercial building, or other.}
- {population_density (created numerical variable): number of which tells us the population over the total area per corresponding neighbourhood group based on ID area_num}

Statistical Approach

We simply used data manipulation and the results of our subquestions to answer this question.

Specifically, we found the population densities of every one of our neighbourhood groups and related them to the mean B & E rates. This data is then translated into a clean summary table providing valuable insights.

Results

area_num	mean_b_and_e_rate	population_density
1	391.7	0.002636601
2	407.8	0.003221365
3	430.4	0.003104700
4	531.3	0.004051301
5	435.5	0.003686601
6	330.4	0.005741494
7	484.4	0.005732440
8	1171.1	0.009007358
9	513.4	0.007989549
10	512.8	0.007478077
11	359.9	0.007637586
12	515.4	0.004552753
13	636.6	0.003701242
14	496.1	0.002777976

Results cont'd

Area 8 has the highest population density, and also the highest overall B & E rate. This finding corresponds to the fact that it has the largest sum of B & E offences to houses and apartments, relating to what we predicted in our first subquestion. It is also reported that area 8 has the highest number of B & E in apartments compared to our 13 other districts.

Other than area 8, area 9, 10 and 11 have high population density, and area 4, 12, and 9 have high B & E rates.

Conclusion/Suggestions

{In a booming city like Toronto, houses and apartments are victimized of break and enter robberies regularly. We want to prevent this.}

{In collaboration with the TPS's data, we conclude our research with three major findings:}

{1. Whether people live in houses or apartments, B & E robberies are likely to happen in either one.} {2. Area number 8, Bay Sreet Corridor, Church-Yonge Corridor, North St. James Town etc. is the hotspot for B & E robberies.} {3. Those who live in houses are quicker to respond to B & E robberies than those living in apartments.}

{We firstly suggest that all people must learn to be prepared to handle B & E robberies, because it can happen to anybody. Then, since area 8 is the most densly populated area, it inherently requires more police for regulation. Interestingly, this area has a comparatively smaller area among our divided areas, meaning it can be more efficiently patroled and reaction times to B & E robberies can be minimized. Finally, we suggest the TPS take action to better educate those who live in apartments against B & E robberies.}