

Statistical Analysis on City of Toronto Break and Enter Reports by Time and Premise Type

How to Efficiently Allocate the Limited Police Resources

Christopher Matthew, Chaerin Song, Peiyuan Wu. Tutorial 0108
Group 5

March 30, 2020

Introduction

We conducted statistical analysis on the dataset of Break and Enter from the Major Crime Indicators (2014 to 2019) occurrences data, offered by the Toronto Police Service.

Our focus is on analyzing the difference in time trend of break and enter crimes between commercial and residence area to suggest more efficient allocation of limited police resources to combat break and enter crimes.

Objectives

- Hypothesis tests on the proportion of occurrences during the night in commercial area and afternoon in residence area
- Linear association between occurrence hour and frequency of break and enter related crimes, specified by premise type: house and apartment (residence) versus commercial.

Data Summary

- We wanted to see if there is any difference in the number of occurrences on different time between residence and commercial buildings, where residence consists of house and apartments, and commercial consists of commercial and others from variable.
- For convenience, we created a variable to count the break and enter occurrences for each .
- We filtered out N/A values from the data and created a variable that divides all premise types into 2 options: residence and commercial.
- Also, we created a variable that divides 24 hour system into 3 different times, each of 8 hours span: Morning, Afternoon, and Night.

Data Summary

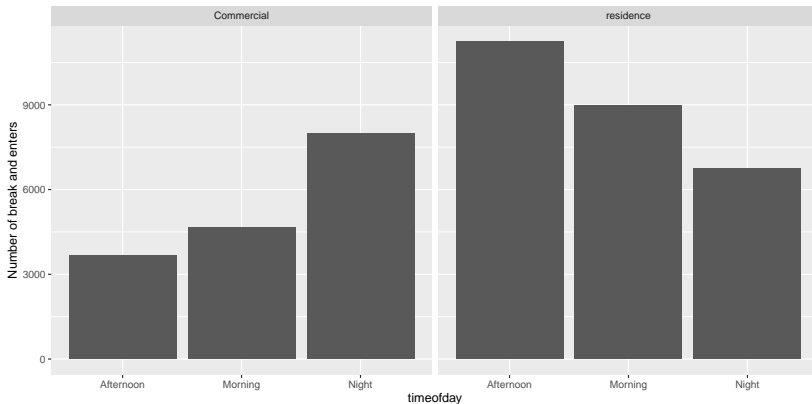


Figure 1: Number of break and enter cases during afternoon, morning, and night occurred in commercial and residence areas

Data Summary

- In figure 1, there is a noticeable difference in occurrence frequency for each time period between residence and commercial buildings.
- This motivated us to run hypothesis testings, each on residence and commercial types, to suggest that there is a difference between the number of break and enter cases among 3 different period in a day, specified by residence and commercial buildings.
- Apparent increase and decrease in the number of break and enter crimes during different time period motivated us to build linear regression model to demonstrate hourly trend (from 10 am to midnight) for residence and commercial type.

Hypothesis Testing

Statistical Methods- Hypothesis

We classified 8 pm to 4 am as Night, 4 am to 12 pm as Morning, and 12 pm to 8 pm as Afternoon, and set our significance level to be 0.05.

Hypothesis (Type: Commercial Area)

- **Null Hypothesis**

$$H_0 : \hat{p}_{commercial,night} = 1/3$$

- **Alternative Hypothesis**

$$H_A : \hat{p}_{commercial,night} \neq 1/3$$

where $\hat{p}_{commercial,night}$ is the proportion of break and enter crimes at night among all break and enter crimes in commercial area.

Statistical Methods- Hypothesis

Hypothesis (Type: residence Area)

- **Null Hypothesis**

$$H_0 : \hat{p}_{\text{residence,afternoon}} = 1/3$$

- **Alternative Hypothesis**

$$H_A : \hat{p}_{\text{residence,afternoon}} \neq 1/3$$

where $\hat{p}_{\text{residence,afternoon}}$ is the the population proportion of break and enter crimes in the afternoon among all break and enter crimes in residence area.

Statistical Methods- Test Statistics

$$p_{commercial,night} = 0.490$$

$$p_{residence,afternoon} = 0.417$$

where $p_{commercial,night}$ is the population proportion of break and enter crimes at night among all break and enter crimes in commercial area, and $p_{residence,afternoon}$ is the population proportion of break and enter crimes in the afternoon among all break and enter crimes in residence area.

Statistical Methods- Sampling Distribution

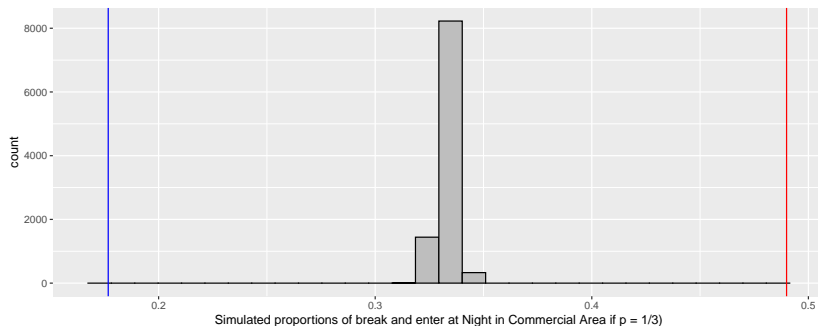


Figure 2: Sampling Distribution of the Proportion of break and enter crimes at Night in Commercial Area

Statistical Methods- Sampling Distribution

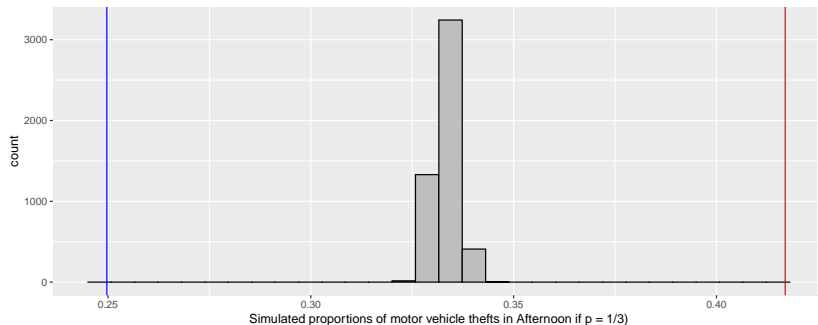


Figure 3: Sampling Distribution of the Proportion of break and enter crimes in the Afternoon in residence Area

Results

Both test statistics for residence and commercial area are way off the range of sampling distributions.

P values for both models are zero.

Model	p-value
residence	0
Commercial	0

Results

Since our p-values for both residence and commercial buildings are zero, there is a very strong evidence against both null hypotheses.

We conclude that it is very likely that there is a difference in the number of break and enter crimes during different time of the day, each in commercial and residence areas.

Linear Regression Models

Statistical Methods

To analyze the daily occurrence trend in residence and commercial building, we filtered the data to only include the crimes that occurred from 10 am to midnight.

We built two linear regression models, each for residence and commercial buildings, to analyze how the number of break and enter cases shifts its trend throughout the day.

Statistical Methods

Hypothesis

For regression model represented by $y_i = \beta_0 + \beta_1 x_i$:

- **Null Hypothesis:** $\beta_1 = 0$
- **Alternative Hypothesis:** $\beta_1 \neq 0$

where y_i and x_i are the number of occurrence the occurrence hour, respectively, β_1 is average change in y for one unit change in x , β_0 is average of y when $x = 0$.

Statistical Methods

Fitting the Linear Regression Models

- We fit the models to find the p-value and conclude whether there is any difference between the number of break and enter cases for different hours or not.
- We fit our linear regression models on (from hour 10 to 24) for each residence and commercial buildings using R.

Results

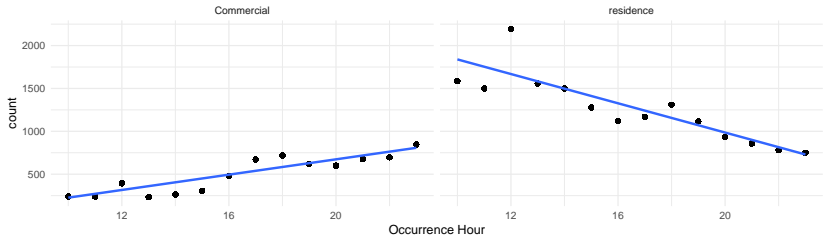


Figure 4: Fitted Regression Models of residence and Commercial Area Occurrence Trend

Results

According to the model, as the day proceeds, there is an increasing trend of cases from commercial buildings while residence type buildings show decrease in cases.

Equations of Fitted Regression Models

- Residence Buildings:

$$\hat{y}_i = 2692 - 85.3x_i$$

- Commercial Buildings:

$$\hat{y}_i = -220.7 + 44.7x_i$$

Note: These only apply for hours from 10 to 24.

Results

P values for both models are zero.

Model	p-value
Residence	0
Commercial	0

Results

Thus, we can conclude that there is a very strong evidence against the null hypothesis.

It is very likely that there is statistically significant difference between the number of break and enter cases throughout the day for both residence and commercial buildings.

According to the model, we conclude that from 10 am to midnight, there is an gradual increase in the number of break and enter crimes in commercial buildings and a decrease in residence buildings.

Conclusion

Finally, we would like to suggest some tips for the Toronto Police Department regarding how to Break and Enter Crimes.

- We can increase safety by educating the public in different premise type on the most dangerous time which most Break and Enter related crimes may happen.
- Expand Neighbourhood watch for Break and Enter related crimes depending on what the major premise type of the buildings are.
- Possibly invest more resources on residence areas during the middle of the day and gradually turn the focus away to commercial areas as the day goes.

Limitations

Even though we have done a lot of statistical inferences, there are still some limitations.

- First of all, single category of premise type could differ its likelihood of break and enter crimes depending on other factors.
- Next, these statistical analysis are just the likelihood, not what is 100 percent accurate.
- Last but not the least, as the years go by, things like technology advances, human capabilities, or city-wide important events can lead to different motives in crimes, which may alter the data or inferences that were made earlier.