Toronto Neighbourhood Crime Rates - Analysis

Using Simple Linear Regression Model

Md Abrar Nasir

February 14, 2020

Introduction

Criminal activities or crime rates have always been a serious point of concern for citizens and authorities in a city. In this analysis, we are to investigate the relationship between the population of a city, which, in our case, is Toronto and its crime rates, based off 2016 data, over varying types of crime categories in question, such as assault, auto theft, robbery etc.

The data for this exploration has been sourced from the Open Data Toronto Portal, published by the Toronto Police Services.

Model

The mathematical model which is to be used is simple linear regression (SLR) which will exhibit the relationship between our response variables, *number of assaults*, *number of auto thefts*, *number of break-and-enters*, *number of homicides*, *number of robberies* and *number of thefts* and the predictor, *population in Toronto in 2016* by means of the linear equation:

$$Y_i = \alpha + \beta x_i + U_i$$

where Y_i is our response variable, x_i is our predictor, and α is the $y-$intercept and β is the slope of the line.

It should also be mentioned that U_i represents any random fluctuation in the relationship, which, for the benefit of simplification, will be assumed to be independent and have an expectation of zero. Also, it should be considered that the variance is same for each U_i. The values of x_i are to be taken as non-random while the values of y_i should be realizations of random variable, Y_i such that (x_i, y_i) is the placeholder for our bivariate dataset containing the regressors and the regressands resprectively.

In layman's terms, this can be explained as drawing a straight line of best fit to express the relationship between our x_i and y_i and then computing the values of the slope and the intercept, to predict crime rates of a certain category using the population number.

To better judge the appropriateness of the model to suit our data, we can examine the scatterplots below which depicts the population sizes and their corresponding crime rates for the set of six types of criminal activites -

As we can see observe in the illustrations, for all, except the `number of homicide ~ population` plot, there is an upward flow of dots, representing a positive trend or relationship. This can be attested by any human rationale suggesting there should be an increase in crime rates (of any type almost) as population increases. Hence, it can be concluded that our SLR model is well-suited with our dataset holding the discrete values.

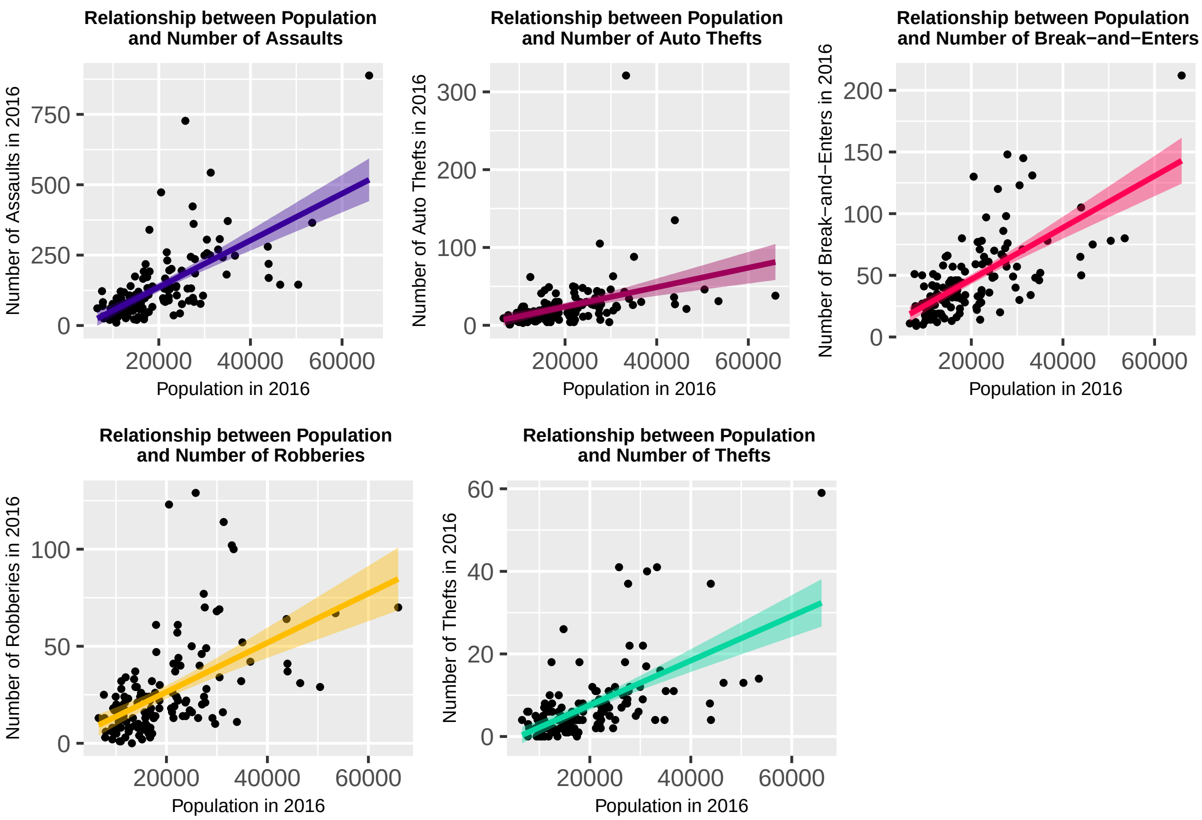
Results

The α and β values for our investigations can be calculated using computer software, which are tabulated below. The `number of homicide ~ population` entry should been dropped from our analysis, as its scatterplot hardly represents a pattern and its corresponding regression line would be unreasonably framed.

Categories	Assault	Auto Theft	Break-and-Enter	Robbery	Theft
α	-29.878	-1.064	4.827	1.062	-3.210
β	0.00831	0.00125	0.00209	0.00126	0.00054

The α value, for each category of crime, shows what the crime rate is supposed to be when the population is zero, as per our model. The β values, on the other hand, show by what proportion would there be a rise in crime rate for each category of crime following a change in population.

In an ideal situation, crime rate should be zero when population is fixed at the zero level. This condition, however, does not align with the information relayed in the table, making the results lose a bit of meaning.



Now that we have the regression lines formulated after the α and β values for the categories of crime in Toronto laid over the scatterplots, it can be easily understood what the relationships are by means of this visualization. The solid line is the line with the model equation $Y_i = \alpha + \beta x_i$ and the shadowed regions represent the confidence intervals for the predicted values.

Conclusion

Finally, in this analysis, we can conclude that there is a positive relationship between neighbourhood crime rates in Toronto and the population in the different neighbourhoods, as suggested by the upward sloping regression lines. It has also been figured out that, while this SLR model is somewhat fitting for the data, there are still some unreasonable attributes to them, such as the case where α values are non-zero.