# A Statistical Study of Crime Rate in Suburban British Columbia

Dawei Sun

The Hill School, Pennsylvania, United States

dsun24@thehill.org

**Abstract.** This study examines the correlation between various factors and crime rates in four regions in British Columbia, Canada. The study focuses on both positive and negative correlators and ranks them based on the strength of their correlation coefficients. The positive correlators include immigrant population percentage, single-parent families, unemployment rate, religious population percentage, and population without employment income. The negative correlators are related to income levels, including the percentage of working population with high annual income, median and average employment income. The study finds that income levels have a strong negative correlation with crime rates, while immigration and certain social structure factors have a positive correlation. However, this study does not establish a causal relationship between these factors and crime rates. The findings suggest the need for further research to understand the specific nature of the relationship between immigration and crime rates, considering contextual factors such as wealth, income level, and employment status of immigrants.

**Keywords:** Crime rate, empirical study, immigration, British Columbia, Canada.

## 1. Introduction

Many methods of crime prevention have been implemented in perhaps every nation in the world, with a large portion of them proving controversial. In this paper, we review several factors commonly associated with high/low crime rates and compare their correlation with actual crime rates across several regions using established statistical methods, therefore determining whether common public perception of the factors associated with crime are correct, while also pinpointing the factors that are good indicators of crime in actuality.

This report first provides a review of recent literature on factors leading to increased or decreased crime. It then gives an overview of the Pearson correlation coefficient and regression analysis, the specific statistical methods used in this study to determine correlation. Next, using those methods, this report provides a statistical analysis of the correlation between several factors, including several common factors from the aforementioned literature review and other general demographic information, and crime rates in four regions in British Columbia, Canada. Lastly, we provide a summary and discussion of the results of the statistical analysis and compare them to the popular beliefs about crime determined in the literature review to gauge their accuracy.

## 2. Literature Review

The question of what factors lead to crime is a much-studied topic. While there is no one factor, or one set of factors, that can predict crime rates with utmost certainty, there are several factors that are generally considered to be very reliable when determining or predicting the crime rates within a certain area.

Crime is, most commonly, linked to immigration. Göran Adamson, a sociologist from Sweden, has linked high crime rates in his country with the influx of migrants in the 21st century [1]. His investigation follows up on a report published by the Swedish Crime Prevention Agency in 2005, which first linked the rising crime rates in Sweden with immigration. His claims are backed up by a study published by Jianxin Cheng, Junqiang Liu, and Jun Wang, a group of Chinese researchers who studied domestic migration within China and found that crime rates were directly related with the percentage of domestic migrants living within rental homes [2].

On the other hand, researchers John H. Boman and Owen Gallupe, who studied decreasing crime rates in the US during the COVID-19 pandemic [3], and Michael Gameli Dziwornu, who studied the crime drop in Ghana [4], conclude that governmental style and response have a large impact on crime as well. Boman and Gallupe concluded that, during the pandemic, governmental response to the situation gave offenders the need to commit graver offenses, while minor crimes usually committed in groups decreased [3]. In Ghana, Dziwornu concluded that the switch from military rule to lawful society caused the corresponding reduction in crime [4]. In the case of the former, it can be inferred that crime rates have a positive correlation with need, while in the case of the latter it can be inferred that social structure can also affect crime rates.

To conclude, the main factors to be most commonly associated with crime rates are immigration, need, and social structure.

## 3. Methodology

Thanks to 2021 census data that was just recently published by Statistics Canada, we were able to obtain crime and population statistics for several regions in Canada [5, 6]. We then used this data to calculate the correlation, across four selected cities in British Columbia, Canada, between the crime rate per 100,000 population (henceforth referred to as crime rate) and several other population statistics. The four regions studied, Surrey, Richmond, Burnaby, and Delta, are chosen due to their similar profiles as suburban cities in the Metro Vancouver Regional District, thus attempting to control for qualitative variables such as culture, climate, geography, political status, etc.

The data sets chosen for analysis include general statistics, such as population density, to the factors that were pinpointed in the literature review, such as immigration population percentage. In this case, income levels are also heavily represented in the sets of data used due to its strong association with "need". Social structure is represented by several factors, such as those relating to education level and family structure. Other statistics that are commonly cited in association with crime rate, such as the unemployment and labour participation rates, are also studied.

The strength of correlation is represented by the Pearson correlation coefficient, a number between -1 and 1 that denotes the strength of correlation between two sets of data points. The formula for deriving this coefficient (r) is:

$$r = \frac{\sum(x_i-\overline{x})\sum(y_i-\overline{y})}{\sqrt{\sum(x_i-\overline{x})^2\sum(y_i-\overline{y})^2}}. \quad (1)$$

Here, x and y represent the two sets being compared, where $x_i$ represents the ith element in x and x represents the mean of all elements in x, with the same being true with $y_i$ and y for the elements in y.

A correlation coefficient of 1 signifies a perfect positive correlation between the two sets, while a correlation coefficient of -1 signifies a perfect negative correlation between the two sets. Generally, a coefficient greater than 0.8 or smaller than -0.8 can be interpreted as a strong positive correlation or strong negative correlation, while a coefficient greater than 0.4 or smaller than -0.4 can be interpreted as a moderate positive correlation or moderate negative correlation, respectively.

The obtained statistics were then input into the formula as the y data set, with the x data set being the crime rate per 100,000 population stat of the corresponding region.

After obtaining the results, we chose the data sets whose absolute correlation with crime rate was greater than 0.4, indicating at least moderate correlation, and also obtained their standard error using regression analysis, which measures a data set's average deviation with the independent set, in this case the crime rate. The standard error of sets with less than moderate correlation with crime rate are not calculated, since the error rate measures the average rate of deviance from the line of best fit and is not meaningful when the dependent variable has little to no correlation with the independent variable.

Here, we use the method of linear regression due to the small number of data points. In regression analysis, linear regression is defined by the equation,

$$y = bx + a + \varepsilon \quad (2)$$

Where x is the independent variable (in this case crime rate), y is the dependent variable that we are comparing it to, a is the y-intercept (the value for y when x reaches 0), b is the slope of the graph (the rate of change of y as a variable of x), and ε is the random error term, which measures the difference between the calculated value on the line of best fit and the actual value from the data set.

## 4. Results

For context, the crime rate for the 4 regions studied are listed below.

**Table 1.** Crime Rate per 100,000 Population in Studied Regions

| Region | Crime Rate per 100,000 Population |
|---|---|
| Surrey | 7,281.62 |
| Richmond | 7,919.36 |
| Burnaby | 5,616.32 |
| Delta | 4,687.04 |

The following results are separated into data sets with a positive correlation with crime rate and data sets with a negative correlation with crime rate, and sorted from strongest to weakest correlation. The standard error is provided as well with the corresponding units which, in most cases, are percentage points (%). All correlation coefficients are round to the nearest thousandth digit, while standard error is round to the nearest tenth digit.

### 4.1. Positive Correlations with Crime Rate

**Table 2.** Positive Correlators with Crime Rate, Ranked by Strength of Correlation with Crime Rate (Highest to Lowest)

| Statistic | Correlation Coefficient with Crime Rate | Standard Error (Units) |
|---|---|---|
| Immigrant Population Percentage | 0.779 | 8.7% |
| Single-Parent Families, as a Percentage of Total Census Families | 0.766 | 1.2% |
| Unemployment Rate* | 0.729 | 1.2% |
| Religious Population Percentage* | 0.719 | 0.2% |
| Single-Parent Families where Parent is Female, as a Percentage of Total Census Families | 0.640 | 2.5% |
| Percentage of Population over 15 without Employment Income* | 0.474 | 2.6% |
| Percentage of Population with no Secondary Education Certificate* | 0.280 | - |
| Population Density per Square Kilometer | 0.263 | - |
| Percentage of Population with a Postsecondary Certificate, Diploma, or Degree* | 0.128 | - |

*25% sample data as provided by Statistics Canada

### 4.2. Negative Correlations with Crime Rate

**Table 3.** Negative Correlators with Crime Rate, Ranked by Strength of Correlation with Crime Rate (Highest to Lowest)

| Statistic | Correlation Coefficient with Crime Rate | Standard Error (Units) |
|---|---|---|
| Percentage of Working Population with Annual Income $125,000 CAD or Above | -0.967 | 0.3% |
| Median Employment Income for Full-Time, Full-Year Workers | -0.905 | 773.8 CAD |
| Average Employment Income for Full-Time, Full-Year Workers | -0.781 | 3595.0 CAD |
| Labour Participation Rate* | -0.396 | - |
| Average Age | -0.234 | - |
| Percentage of Population to have Never Married | -0.084 | - |
| Percentage of Population Married or Living Common Law | -0.075 | - |
| Percentage of One-Person Households | -0.005 | - |

*25% sample data as provided by Statistics Canada

## 5. Conclusion

### 5.1. Overview of Results

In this analysis, we found six positive correlators with crime rate whose strength of correlation was at least moderate (correlation coefficient > 0.4). They were, in order of strength of correlation:

1) Immigrant Population Percentage;
2) Single-Parent Families, as a Percentage of Total Census Families;
3) Unemployment Rate;
4) Religious Population Percentage;
5) Single-Parent Families where Parent is Female, as a Percentage of Total Census Families;
6) Percentage of Population over 15 without Employment Income.

Of these six categories, Religious Population Percentage had the smallest standard error, with just a 0.2% deviation. This is likely due to the high percentage of the population that is religious in all these regions, reducing the margin of error rate. The category with the greatest strength of correlation with crime rate, Immigrant Population Percentage, had an outstanding error rate of 8.7%. This may be due to the highly contextual nature of immigration and the corresponding difficulty in generalizing all immigrants in one category.

Additionally, we found three negative correlators with moderate or greater correlation with crime rate (correlation coefficient < -0.4). Those, in order of strength of correlation, were:

1) Percentage of Working Population with Annual Income $125,000 CAD or above;
2) Median Employment Income for Full-Time, Full-Year Workers;
3) Average Employment Income for Full-Time, Full-Year Workers.

Notably, all three correlators were measures of income level. Interestingly, there is quite a large gap between median employment income and average employment income in terms of both their correlation coefficients and error rates.

### 5.2. Discussion of Findings

While there are fewer negative correlators studied in this study that have an at least moderate correlation with crime rate, the strongest correlators tended to be negative, with the 3 strongest correlators studied all being negative. These 3 factors were all financial, showing the strong relationship between crime and wealth. A worthwhile follow-up to this study might study this

relationship further to determine whether there is a cause-and-effect relationship between crime and wealth. Interestingly, there was a significant deviance between the correlation coefficient of the median and average incomes (-0.905 vs. -0.781) and also the error rate between the two (773.82 CAD vs. 3594.96 CAD). This can be accounted for by the difference in wealth distribution between the four regions.

Nevertheless, the positive correlators are still worth examining. The strongest positive correlation identified in the study was Immigration Population Percentage, with a correlation coefficient with 0.779. However, the standard error rate in this case is outstanding, with a standard error of 8.7 percentage points. This is significantly greater than the error rate of any other category studied, even including those with a weaker correlation. The percentage-based data set with the second greatest error rate, Percentage of Population over 15 without Employment Income, had an error rate of just 2.6 percentage points, less than ⅓ that of Immigration Population Percentage, despite having a correlation coefficient of just 0.474. This, as discussed previously, may be explained by the highly contextual nature of immigration, especially in an immigration-heavy region such as British Columbia. The four regions examined have widely varying places of origin for their immigrant population; for example, the percentage of immigrants born in China for Surrey, Richmond, Burnaby, and Delta are 11.1%, 51.7%, 25.4%, and 12.5%, respectively. A follow-up study may examine immigration's effect on, or from, crime rate with the specific context in mind.




**Figure 1.** Country of Origin of Immigrant Population An illustration of the difference between the four regions' immigration population makeup

Other strong correlators included single-parent family percentage, unemployment levels, religious population percentage, and education level.

In conclusion, the negative correlators found in this study show the strong correlation that crime rate has with income levels. On the other hand, the strongest positive indicators include immigration, number of single-parent families, and unemployment rate.

### 5.3. Comparison with Previous Results

These results corroborate Göran's and Cheng, Liu, and Wang's claims that immigration can be a relatively strong indicator of crime, as immigration rate had the strongest positive correlation with crime rate of any variable [1, 2]. However, their more specific conclusions were region-dependent and cannot be supported by our conclusions. The strong correlation between income levels and crime rate also heavily support Boman and Gallupe's studies of need-based crime [3]. The somewhat strong positive correlation of several factors indicative of social structure, such as unemployment and education, lend some credence to Dziwornu's claims that social structure affects crime rates [4].

Specifically, the positive correlation between education levels, which are highly indicative of lawful society, and crime rate heavily supports his claim that less crimes are committed in a lawful society.

### 5.4. Limitations and Further Studies

One limitation of this study is its small scale, with only four regions studied and compared in both our correlation tests and regressions analysis tests. This was done intentionally in order to control for the qualitative differences between regions, such as geography, culture, political status, etc., thus studying the selected factors in four geographically and culturally similar regions. However, it is possible that a sufficiently large sample size of regions will also control for these factors, as a sufficiently large sample size would likely average these factors out to the mean.

Because this study focuses on identifying factors that correlate with crime, another limitation of this study is that our conclusions cannot be used to determine whether these factors are causes or effects of crime. Because of this, future studies may focus on whether these factors, such as income levels and immigration rate, affect crime rates more or less than crime rates affect them.

Furthermore, this study found that further examination of immigration's relationship with crime is needed. While this study and several others, such as those by Göran and Cheng, Liu, and Wang, conclude that there is a positive correlation between crime rate and immigration rate, the exact nature of that relationship is not known [1, 2]. The high error rate found in this study shows that, while studying immigration, further context is needed. Considering the other results found within this study, the wealth, income level, and employment status of incoming immigrants may be some factors that should be taken into account in future studies.

### 5.5. Summary

In this study, we concluded that the factors identified in the literature review, namely immigration, need, and social structure, are indeed factors that are correlated with crime rate. Income level statistics are the most indicative of the metrics studied, whereas immigration population percentage is another highly indicative metric. However, its high error rate implies that it is a highly contextual statistic that is difficult to make inferences from. Other strong correlators included single-parent family percentage, unemployment levels, religious population percentage, and education level.

This study was not able to determine the cause-and-effect relationship between crime rate and any of the examined statistics. Thus, the nature of their relationship with crime rate should be an area of interest for future studies.

## Acknowledgements

I, Dawei Sun, hereby declare that the work presented in this paper is my own original work. Where information has been derived from other sources, I confirm that this has been clearly and fully identified and acknowledged.

## References

[1] G. Adamson, "Migrants and Crime in Sweden in the Twenty-First Century," Society, vol. 57, no. 1, pp. 9–21, 2020, Doi: 10.1007/s12115-019-00436-8.

[2] J. Cheng, J. Liu, and J. Wang, "Domestic migration, home rentals, and crime rates in China," The Journal of Chinese Sociology, vol. 4, no. 1, p. 8, 2017, Doi: 10.1186/s40711-017-0056-3.

[3] J. H. Boman and O. Gallupe, "Has COVID-19 Changed Crime? Crime Rates in the United States during the Pandemic," American Journal of Criminal Justice, vol.45, no.4, pp.537–545, 2020, Doi: 10.1007/s12103-020-09551-3.

[4] M. G. Dziwornu, "Crime drop in Ghana? Some insights from crime patterns and trends," Crime Prevention and Community Safety, vol. 23, no.4, pp. 433–449, 2021, Doi: 10.1057/s41300-021-00130-0.

[5] Statistics Canada, "Census Profile, 2021 Census of Population," Government of Canada, Mar. 2023.

[6] "Table 35-10-0184-01 Incident-based crime statistics, by detailed violations, police services in British Columbia," Statistics Canada, Sep. 21, 2023.