

For Assignment 3 – Part 1 and Part 2

Statement of Academic Integrity: This assignment conforms to the rules on academic integrity of the University of Ottawa: Name in place of Signature: Shahid Jadavji (6037549)

Assignment 3 – Part 2 – Question 1

Crime in Canada

Is crime worse in larger cities? Many people think it is, but what does the data say? Of course there are many types of crime, some worse than others. A large city might have more traffic offenses and less assault cases. We need a way of combining all types of crime together, weighted according to how severe the crime is. That is what Statistics Canada's "crime severity index" does. Each type of offense is assigned a weight derived from actual sentences handed down by courts. More serious crimes are assigned higher weights, less serious offences lower weights. As a result, the index reflects the overall severity of crime in a given city. For a full explanation of the index and how it is calculated, you can refer to the report "measuring Crime in Canada: Introducing the Crime Severity Index and improvements to the Uniform Crime Reporting Survey" (Statistics Canada publication number 85-004-X).

The data file Ch07_MCSP_Crime_in_Canada contains the crime severity index and the population of certain Canadian cities. Draw scatter plots and calculate the correlation coefficients for the whole of Canada and then separately for each region of Canada. (Ottawa-Gatineau should be included in both the Ontario and Quebec regions.) Comment on whether you think your correlation coefficients are representative of those regions, giving reasons for any doubts you may have. Finally answer the million dollar question: Is crime worse in larger cities?

Use the PLAN/DO/REPORT format to answer this question.

PLAN

Using the data found in the study conducted by Statistics Canada labelled “*Measuring Crime in Canada: Introducing the Crime Severity Index and Improvements to the Uniform Crime Reporting Survey (Statistics Canada Publication 85-004-X)*” we attempt to approximate an answer to the question “Is crime worse in larger cities?”. The data provides the population for each region with its corresponding crime severity index value with respects to the region that the city is situated in.

The quantitative variable “crime severity index” is measured through a value found in the index.

The quantitative variable “population” is a measure approximated with respects to its region.

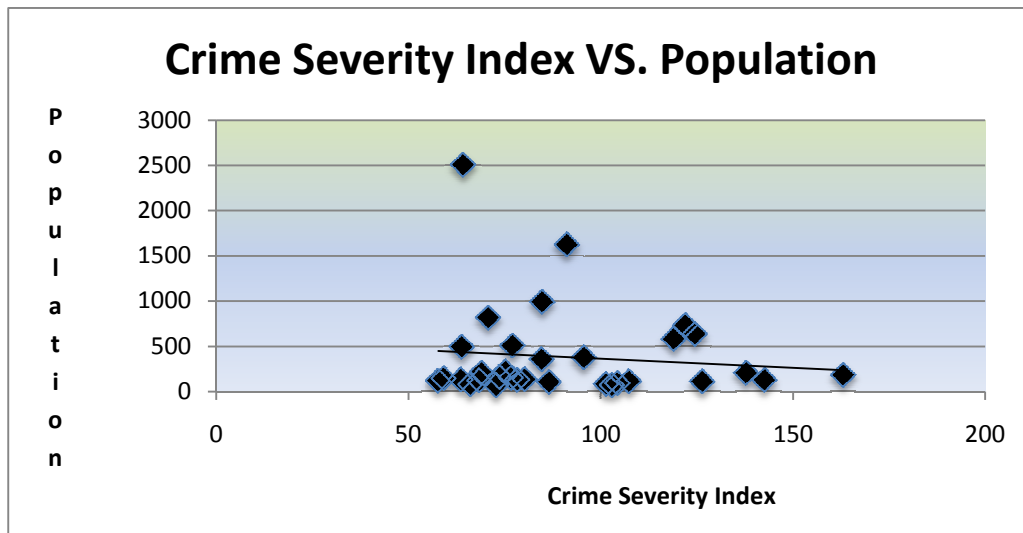
We use statistical tools like scatter plots and the correlation coefficient as a function for each region to determine the extent of relatedness between the variables.

The scatter plots and correlation coefficient for each region can be seen below in the DO portion of this statistical analysis. The data provided can be seen below.

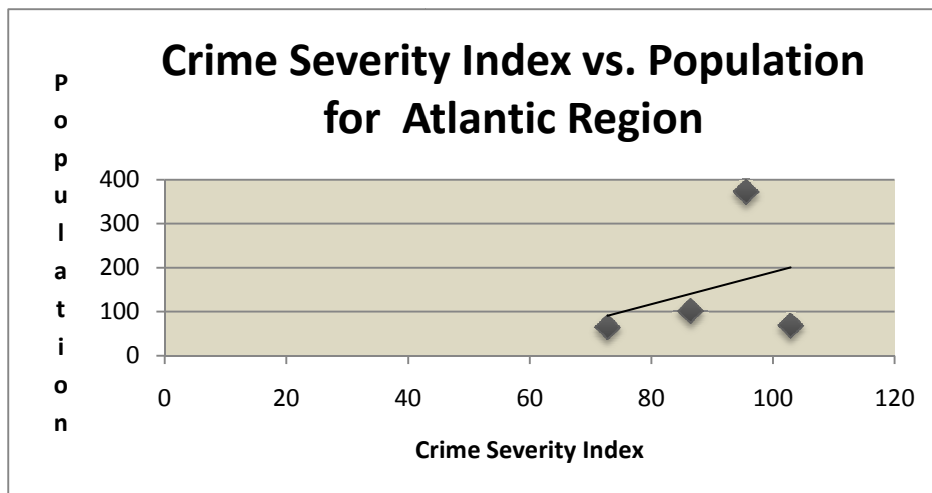
DO

City	Region	Crime Severity Index	Population (thousands)
St. John's, Newfoundland and Labrador	Atlantic	86.44	101
Halifax, Nova Scotia	Atlantic	95.58	372
Saint John, New Brunswick	Atlantic	102.9	68
Moncton, New Brunswick	Atlantic	72.79	64
Saguenay, Quebec	Quebec	59.23	143
Québec, Quebec	Quebec	63.82	491
Sherbrooke, Quebec	Quebec	76.62	147
Trois-Rivières, Quebec	Quebec	78.23	126
Montréal, Quebec	Quebec	91.16	1621
Ottawa-Gatineau, Ontario/Quebec	Ontario and Quebec	70.69	812
Toronto, Ontario	Ontario	64.17	2503
Hamilton, Ontario	Ontario	77.04	505
St.Catharines-Niagara, Ontario	Ontario	80.13	131
Kitchener, Ontario	Ontario	68.92	205
London, Ontario	Ontario	84.57	352
Windsor, Ontario	Ontario	75.28	216
Sudbury, Ontario	Ontario	73.63	158
Thunder Bay, Ontario	Ontario	107.29	109
Kingston, Ontario	Ontario	68.42	117
Peterborough, Ontario	Ontario	66.14	75
Brantford, Ontario	Ontario	104.34	90
Guelph, Ontario	Ontario	57.74	115
Barrie, Ontario	Ontario	63.61	128
Winnipeg, Manitoba	Prairies	124.42	633

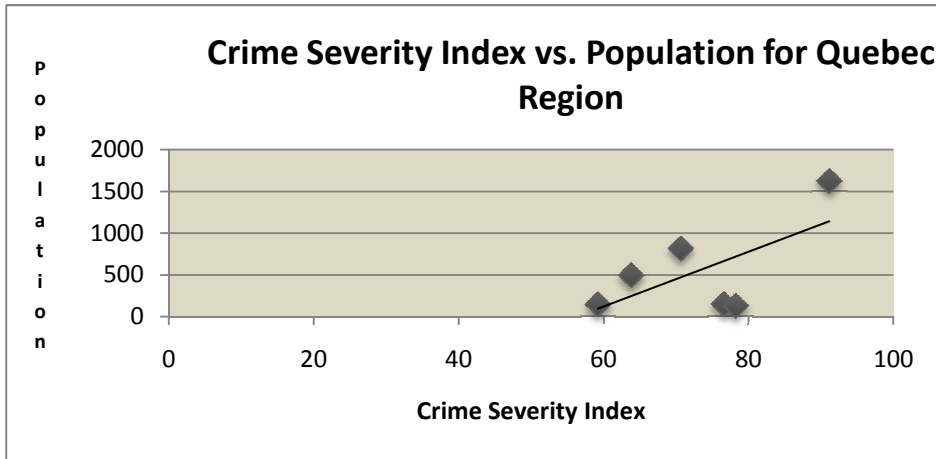
Regina, Saskatchewan	Prairies	163.05	179
Saskatoon, Saskatchewan	Prairies	137.82	202
Calgary, Alberta	West	84.72	988
Edmonton, Alberta	West	121.99	730
Vancouver, British Columbia	West	119	578
Victoria, British Columbia	West	101.42	78
Abbotsford-Mission, British Columbia	West	142.52	124
Kelowna, British Columbia	West	126.39	107



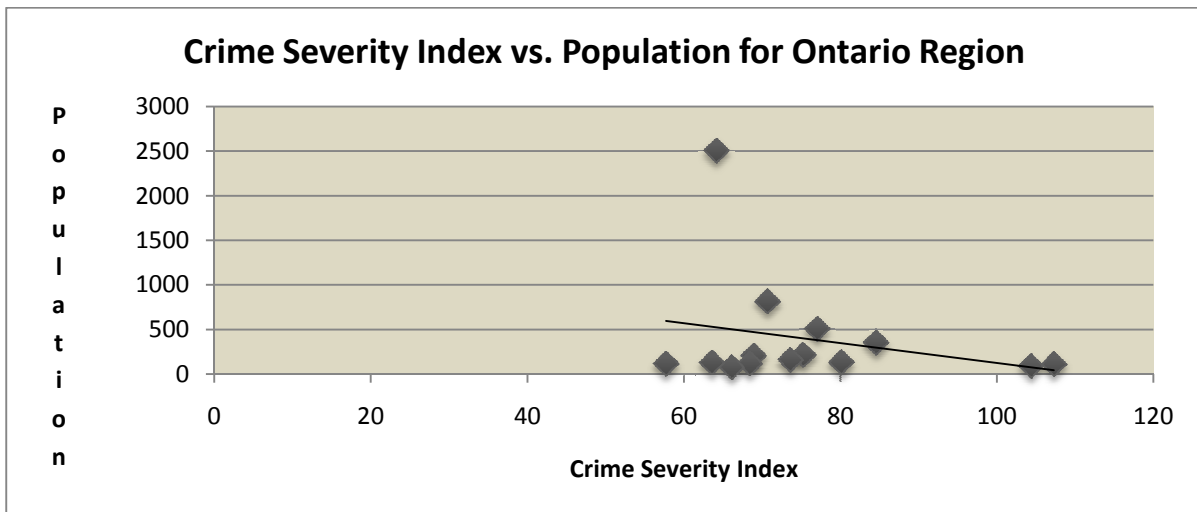
-0.107215943 Correlation for all of Canada



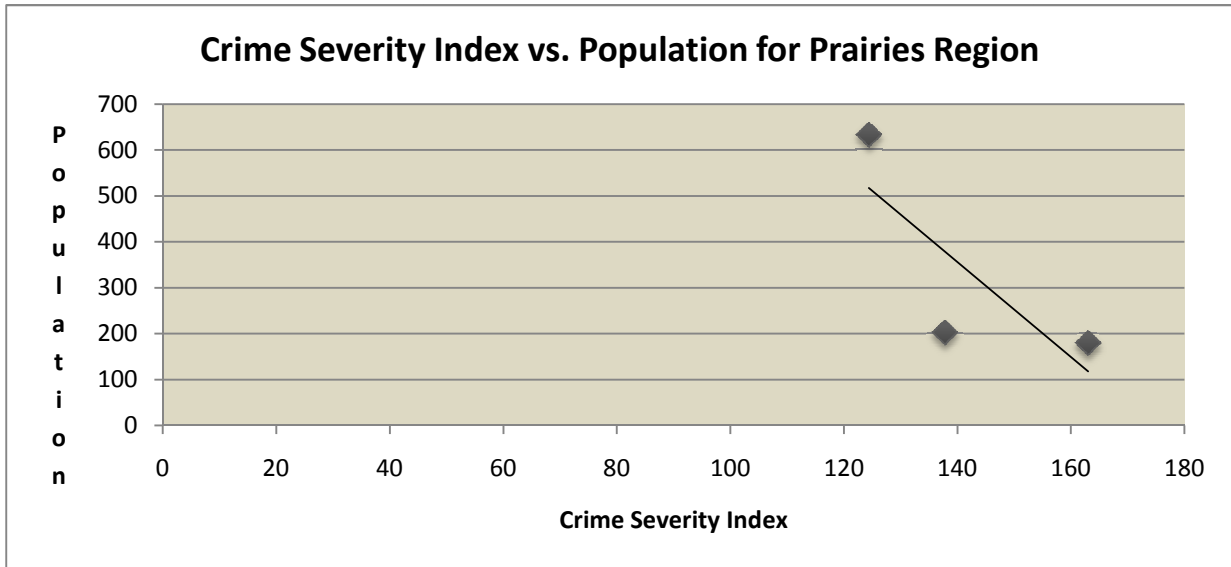
0.318879921 Correlation for Atlantic Region



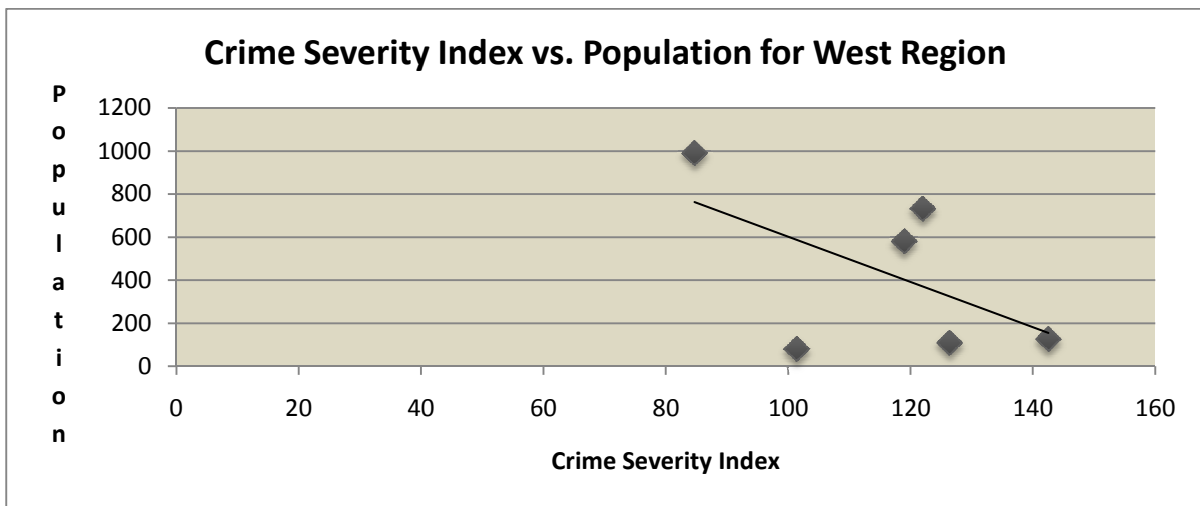
0.636360087 Correlation for Quebec Region



-0.2513521 Correlation for Ontario Region



-0.793891579 Correlation for Prairies Region



-0.551862732 Correlation for West Region

Here, the trend lines help by providing a visual representation for any relationship between the two quantitative variables if they exist. However, the presence of outliers can have a significant impact on the correlation coefficient. After examining the data, its safe to say that there is no presence of obvious outliers.

Report

As a reminder, we are trying to answer the question “Is crime worse in larger cities”.

After analyzing the correlation coefficients for each region, the evidence in the data proves that crime is NOT worse in cities with greater populations. In fact, Quebec, with the correlation coefficient of 0.63636 is the only moderate correlation when compared to the other respective correlation for other regions. The other regions do not have a correlation coefficient that is close enough to 1 to assume that as population increases, the crime severity index increases as well.

Using correlation coefficients as regional representations is beneficial as it provides a quantitative value on the strength of the relationship between crime severity index and population. An imperfection of using correlation coefficients in this exercise is credited to regions such as the Prairies and Atlantic where there is minimal data (3 and 4 records respectively). The sample size is tiny compared to other regions; thus, may resulting in a misrepresentation of data from using correlation coefficients.

Assignment 3 – Part 2 – Question 2

As a real-estate agent, you rely on commissions from house sales and rentals. The number of houses that you sell per year is a random variable (denote X). The number of rental agreements that you secure per year is also a random variable (denote Y).

a) [1] Do you think the random variables X (house sales) and Y (house rentals) are correlated? Briefly explain, making sure to explain the term correlation.

The joint probability models for X (bottom row) and Y (last column).

House Sales/year (X)

House Sales/year (X)

1 3 6 $P(Y=y)$

Rentals/yr (Y)

1 0 0.03 0.12 0.15

4 0.02 0.12 0.11 0.25

10 0.13 0.44 0.03 0.6

$P(X=x)$ 0.15 0.59 0.26

b) [2.5] Compute the following for EITHER random variable X or Y (State which RV you have chosen). Report your answers to 3 significant digits.

i) expected value

ii) variance

Now for the remainder of this question (parts c-d) disregard your answers from above and instead assume the following tabulation of results (if and when you need them). You may assume that the negative correlation between x and y is linear.

$E[X]$ $E[Y]$ $SD[X]$ $SD[Y]$ Correlation(X,Y)

4 sales/yr 8 rentals/yr 2 sales/yr 5 rentals/yr -0.40

c) [5] Assume a fixed commission for each house sale of \$15000/sale and a fixed commission for each rental of \$2500/rental. Determine the following for your annual gross (before tax) income:

i) expected value

ii) variance

d) [1.5] Assume that your annual gross income is Normally distributed. Specify the upper quartile, Q_3 , for your annual gross income.

A) What is Correlation?

- ✓ The correlation ranges from 1.0 to -1.0 where -1.0 is the strongest negative linear relationship, 1.0 is the strongest positive linear relationship and 0 is no linear relationship present.
- ✓ A correlation is a measure of strength pertaining to the relationship between two variables.
- ✓ Correlation can be quantitatively measured using the correlation coefficient

I believe that the random variables X (house sales) and Y (house rentals) are not correlated.

After examining the above distribution function it can be seen that the data is not consistent with continuous distribution functions. Thus, selling a home and receiving commission cannot ensure securing a rental as well.

My Thoughts: In reality these variables MUST be related and dependant because in general significant increase in one RV (X or Y) can lead to a decrease in available houses for sale or for rent. Thus, if there is one more renter(Y), it leaves one less house for someone to BUY (X).

B) Random Variable (X):

Expected Value (X) = [House sales/year * (P(x) for that year)]

$$= [1*(0.15)] + [3*(0.59)] + [6*(0.26)]$$

$$E(X) = 3.48$$

Variance (X) = $\sum [x-E(X)]^2 * P(x)$

$$= [(1 - 3.48)^2] * 0.15 + [(3 - 3.48)^2] * 0.59 + [(6 - 3.48)^2] * 0.26$$

$$= 0.92256 + 0.135936 + 1.651104$$

$$\text{Var}(X) = 2.7096$$

C) Random Variable (Y)

Expected Value

$$E(XY) = 4*(15000) + 8*(2500)$$

$$= 60,000 + 20,000$$

$$E(XY) = 80,000$$

Variance

$$\text{Var}(aX) = (15000^2)*4 = 900,000,000$$

$$\text{Var}(aY) = (2500^2)*8 = 50,000,000$$

$$\text{Var} = 900,000,000 + 50,000,000 + [2*(1500)*(2500)*(-0.4)*(2)*(5)]$$

$$\text{Var}(aX + aY) = 920,000,000$$

D) Upper Quartile

$$Q3 = 0.75$$

Through looking at the Z- Score table for the Normal distribution (continuous) we find that the closest Z-score to 0.75 is 0.68.

$$Z = 0.68$$

$$Z = X - E(x) / SD(x)$$

$$0.68 = x - 80000 / \sqrt{920,000,000}$$

$$X = 0.68 * (30,331.5) + 80,000$$

$$X = 100,625.42$$

The upper quartile, Q3, is the level of gross income, so that there is a probability of 0.25 of getting a gross income higher than that level, X.

For Assignment 3 – Part 1 and Part 2

Statement of Academic Integrity: This assignment conforms to the rules on academic integrity of the University of Ottawa: Name in place of Signature: Shahid Jadavji (6037549)