

CSC110 Project: Impact of COVID-19 on Unemployment Rate and Crime in Canada

Chen Jiang, Kunping Quan, Jiahao Gu, Yufei Liu

Friday, December 14, 2021

Introduction

Overview of background: The pandemic has led to many changes in our daily life. There are fewer job opportunities and people tend to stay at home more. We can see from the (*Canada Unemployment Rate | 2021 Data | 2022 Forecast | 1966-2020 Historical | Calendar, n.d.*) and (Government of Canada, 2021) data set that Canada experienced a significant increase of unemployment rate in 2020, during the breakout of the pandemic. And according to the (*Statistical Relationships between Crime Trends and Major Socio-Economic Trends, n.d.*) and (Government of Canada, 2021b) analysis, when the unemployment rate varies by 1%, the growth rate in homicides varies by approximately 0.39% in the same direction. So, there may be some correlation between the pandemic and the unemployment rate, which is positively associated with crimes. However, because of the transmissibility of the coronavirus, people prefer to stay at home rather than going outside. Thus, fewer people on the street can possibly lead to a decrease in crime since there are fewer people that can be aggrieved in the public. Under such conditions, it's worth discussing about what kind of impact COVID-19 will have on the unemployment rate and crime rate in Canada.

Context for the problem & Motivation: There are many crimes such as robbery and theft happening everyday, which severely jeopardizes our safety. Additionally, we observed that more and more people are wandering on the street. A couple of days ago, one of our friends was robbed of a Canada Goose down jacket in the street. They didn't ask for money but for a down jacket of a famous brand. This implies that there are various reasons behind a crime. Especially under the background of pandemic, the reasons can be complex and diverse. What's more, many people lost their jobs because of the pandemic. Sometimes people who lost their jobs might do extreme things such as committing a crime in order to earn money to survive, or take revenge on the society. So, we want to find out the effects of the pandemic on Canada's unemployment rate, criminal rate and the relationship between the two.

As a result, we have the research question about "**How has the pandemic impact on unemployment rate and the criminal rate, and whether there is a relationship between unemployment rate and crime rate in Canada?**".

There are also some sub-questions that we are going to discuss about:

1. How has pandemic impact unemployment rate?
2. How has pandemic impact the number of crime committed?
3. How has pandemic impact the number of crimes related to money committed?
4. Is there a relationship between unemployment rate and total crime committed?
5. Is there a relationship between unemployment rate and total crime committed?

Dataset Description

For our problem, we have gathered two related data sets that could potentially explain the pattern we are trying to explore. One of them contains the monthly employment information of Canada, and the other one contains the reported crimes information during the Covid-19 pandemic.

DATA1 Employment and unemployment rate, monthly, unadjusted for seasonality

- Source: Statistics Canada official website (<https://www.statcan.gc.ca/>)

- Format: CSV
- Data used in this file: Unemployment rate, Total, all population centres and rural areas
- Description: This data set contain two main variables, the number of employments and the unemployment rate of Canada during a specific time period that is breakdown into observations of months. For both of the main variables, the employment number and unemployment rate are further categorized based on geographic area, which includes core, fringe and rural areas. In conclusion, this data contains the monthly values of employment and the monthly unemployment rates of different areas in Canada.

DATA2 Selected police-reported crime and calls for service during the COVID-19 pandemic

- Source: Statistics Canada official website (<https://www.statcan.gc.ca/>)
- Format: CSV
- Data used in this file: Geography category, total crime number is the sum of all crime categories except total assaults. money related number is the sum of crime categories Total robbery, Motor vehicle theft and Shoplifting.
- Description: This data set contains only one main variables, the number of different kind of crimes reported by the police, and it is shown in a monthly observation during the covid period. One characteristic of this data is that it divides crimes into specific categories, which is especially helpful for our research since we can discuss the relationship between the unemployment rate and a specific kind of crimes (e.g crimes that involves money), and possibly help us to explain the general impact of the pandemic.

Computational Overview

data_conversion.py: For this project, we would only need part of the data set, thus, for efficiency and convenience, we filtered out unnecessary data, and convert the csv data file into Python data types.

- First, we create a dataclass called *Data* including attributes that are related to our research problem, this class combine all information needed in the two separate datasets. What each attribute represents is stated in the Python file.
- We then use *csv* library to import the data file, and convert them into a list of tuples, each tuple contain information of a specific variable of the original file.
- Lastly, we filter out unwanted data in the list we produce above, and store them into the dataclass we created. For example, for the analysis of crimes related to money, we only pick out variables Total robbery, Motor vehicle theft and Shoplifting. Similar operations are implemented to filter out needed data, and is stored into a new list *cleaned_data*, as well as a function *clean_data* to return this cleaned list for future use.

linear_regression.py: For this part, we create the linear regression model for the data, we computed the slope, intercept, and make future predictions for each sub question listed in the introduction.

- First, we created linear regression model for each question by using *numpy* and *sklearn.linear_model* libraries. We used *numpy.array* to make x and y axis variables. Then we created our model by using *LinearRegression()* function from *sklearn* library. And we used *model.fit()* function to fit our x value and y value. Then we used *model.score* to calculate the coefficient of determination, we used *model.intercept_* to calculate the intercept of the equation, and used *model.coef_* to calculate the slope of our equation. There are totally five relationships we have covered: 1. the relationship between COVID-19 and total crime number, 2. the relationship between COVID-19 and money related crime number, 3. the relationship between COVID-19 and unemployment rate, 4. the relationship between unemployment rate and total crime number, 5. the relationship between unemployment rate and money related crime number. We have used the linear regression model to compute the slope and intercept of each relationship.
- Second, we have created five scatter plots by using the *matplotlib.pyplot* library and fitted a linear regression line to each plot, which correspond to each sub question listed in the first part. We used *plt.scatter* to draw the scatter plot between dependent and independent variables, and used *plt.plot* function to fit in our model. Then we labeled x and y axis by using *plt.xlabel* and *plt.ylabel* separately. Finally we used *plt.show()* function to show our final graph. That visualizes the relationship between each two categorization.

- Lastly, we used the linear regression model we have computed to predict the future crime number, money related crime number and unemployment rate based on the certain period that user input. Users input specific month and year, and the linear regression model return the predicted value calculated based on the month and year.

randomization_test.py: Knowing that the null hypothesis is that there is no difference in the mean of unemployment rate for months with and without COVID-19. We do the following steps to apply randomization test to Apply randomization test to the mean of Canadian unemployment rate before and after COVID-19 breakout in Canada (January 2020). For this part, we use *matplotlib* library to produce the distribution of the randomization test.

- Firstly, we calculate the test statistic from observed data. We do that by separating all the unemployment rates into two lists. One with all unemployment rates before COVID-19 breakout and one with all unemployment rates after COVID-19 breakout. Then, we calculate the means of unemployment rate before and after COVID-19 breakout. We get the test statistic by subtracting the two means.
- Secondly, we simulate sample assuming null hypothesis is true and calculate the statistic for each sample. We want to do 1000 simulations. First, create a list of 1000 empty strings. Second, we randomly shuffle the cleaned data into two groups. Since the original data has 10 elements with month before COVID-19 breakout and 21 elements after COVID-19 breakout, group one has 21 elements, and we assume they are after COVID-19 breakout. Group two has 10 elements, and we assume they are before COVID-19 breakout. Third, we calculate the means of unemployment rate of the two groups. Fourth, we calculate the difference of means of the two groups and store it by mutating an empty string in the list created in the first step. Fifth, repeat step two to four 1000 times using a for loop.
- Thirdly, our program reports the results of your computation by visualizing the graph of the sampling distribution base on the list we got in step two, which contains 1000 simulated mean difference of unemployment rate before and after COVID-19 breakout. We use new library called *matplotlib.pyplot* to accomplish this. We use *plt.hist* to draw the sampling distribution as a histogram, we use *plt.axvline* to draw two vertical lines representing the test statistics. Then we use *plt.xlabel*, *plt.ylabel* to add labels for x-axis and y-axis. Finally, we use *plt.title* to add title for the sampling distribution.
- Fourthly, we calculate the p-value. P-value is the possibility of getting unusual data assuming null hypothesis is true. We calculate it by dividing the number of simulations with the absolute value of mean difference larger than the absolute value of test statistic by the total number of simulations, which is 1000, to get the p-value.
- The procedure is the same with the randomization test of number of total crime and number of money-related crime, we just need to change the observation to be the number of total crime and the number of money-related crime.

main.py: this code mainly functions as a summary for the above analysis, which generates interactive windows for viewers to access the results of our computation. There is not much computation in this part, as it is mostly importing results from the above analysis. The *tkinter* library is used here to create windows and the results are imported from the above files.

- The main function of this file is *open_window()*, which contains methods of *tkinter* that create wanted window. Since we want to create two windows at the same time, and we can only have one window mainloop, inside the outer function *open_window()* there is another function *run_prediction_window()*, which would create the prediction window. By calling the *open_window()*, both of the windows could be generated at the same time.
- The functions we used from the library *tkinter* are:
 1. *tk.Tk()*, *window.title()*, *window.geometry()*, *window.configure()*. They are used to generate the window and modify its configures with inputs about its title, size, and background colour.
 2. *tk.StringVar()* and *var.set()*. *tk.StringVar()* is used to create a string variable to be shown up in the labels and *var.set()* is used to modify the value of this variable based on the different outcomes of the different inputs of the user.
 3. *tk.Label().place()* and *tk.Button().place()*. They are used to generate different labels and buttons that have different roles. The input parameters are our different demands of their name, size, colors, and effects. The *.place()* function has input parameters of their positions, determined by the x-coordinates and y-coordinates. They don't return anything.

- In the coding above, we have defined functions that generate visualizations or results of our analysis, and in this coding we import and directly call those functions to give the buttons interactive functionalities.
- There is another function in this code called *show_linear_regression_models()*, which essentially generate interactive buttons for the linear regression part. This function is created only for the code to be in a clearer way.
- As the user clicks the buttons shown on the “Visualization Window”, the corresponding visualizations built by our functions in the *linear_regression.py* and *randomization_test.py* with the p-values as described above will show up. As the year and month the user wants to predict about are typed into the “Prediction Window”, the corresponding values of our predicted values of future total crimes, future money crimes, and future unemployment rate based on the prediction model we built in *textsllinear_regression.py* will show up.

Instructions for obtaining data sets and running the program

Data downloading URLs:

- Crime data: https://drive.google.com/file/d/1nk_6faG86Lv2_SZEo4oMwZTsVnT9v0HR/view?usp=sharing
- Unemployment data: <https://drive.google.com/file/d/1WWxaNlSmTRzzp0PdcTu02u2q7RIBr9rN/view?usp=sharing>
- Please note: the URL link we have provided is the download link from the government Canada. However, it is possible that the URL link does not work as the website sometimes goes wrong under the large clicks. However, we have included the data set in our Markus submission. If you cannot download the URL link, please check our Markus submission. And also, we have mutated the data set file name within our submission and code in order for us to use the data set more conveniently..

Download the above data file and save them to the same folder where other Python files are located. If the name of the files is not as stated below, please change the Unemployment data file name to “unemploymentrate.csv”, and change the Crime rate data file name to “crimerate.csv” for the program to run.

Running program:

To check the result of our test, you need to run the file `main.py` in python console, and type `show_window()` in the console, and the two windows below will be shown.

Prediction window

Prediction Window

The predicted value of future total crime: 39685.43

The predicted value of future money crime: 8643.64

The predicted value of future unemployment rate: 10.1

Please choose the year and month you want to predict about here and click "Generate Output":

Month: 1 2 3 4 5 6 7 8 9 10 11 12

Year: 2019 2020 2021 2022 2023 2024

Generate output

This window make prediction for future total crime number, money related crime number and unemployment rate based on the linear regression model of our analysis.

To make prediction, select the month and the year you want to make prediction on, and press "Generate output" bottom on the left, and results will be shown above in three corresponding white box.

- Note that to make reasonable prediction, please do not select year and month before March 2019, as the data contains only information after this time point.

Visualization window

Visualization Window

Linear regression model of covid and total crime number

Linear regression model of covid and money crime number

Linear regression model of covid and unemployment rate

Linear regression model of unemployment rate and crime

Linear regression model of unemployment rate and money crime

Null Hypothesis: There is no difference in the mean of number of total crime for months with and without COVID-19.

Mean difference of number of total crime

P-value: 0.0

Null Hypothesis: There is no difference in the mean of money-related crime for months with and without COVID-19.

Mean difference of money-related crime

P-value: 0.0

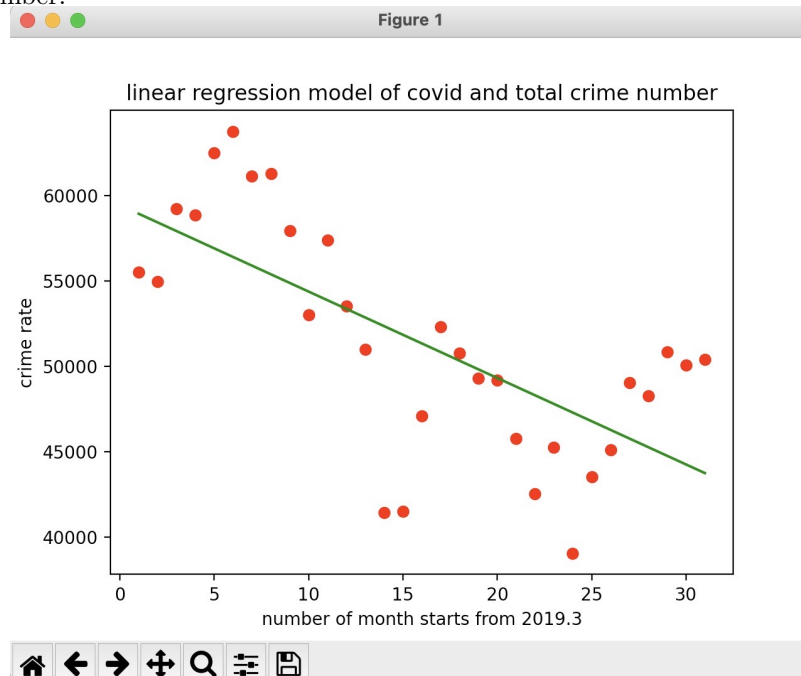
Null Hypothesis: There is no difference in the mean of unemployment rate for months with and without COVID-19.

Mean difference of unemployment rate

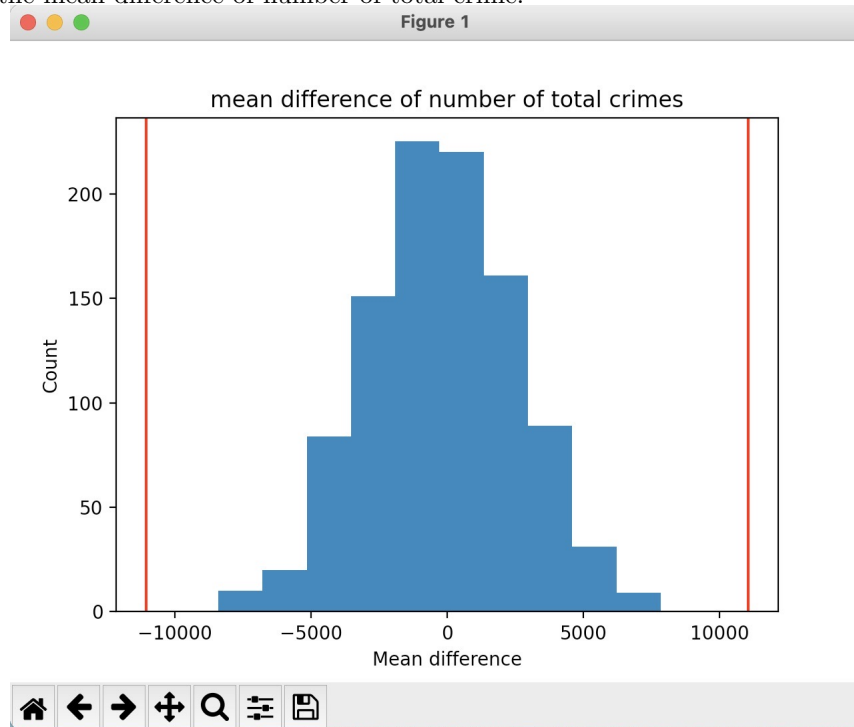
P-value: 0.001

This window will show all analytical results and visualization of our code, including the linear regression graphs, the distribution of our hypothesis test, and the p-value of the test.

To check the visualizations of linear regression models, choose one of the top five bottoms as their title described, the below screenshot is an example of the first bottom, which is the linear regression model of covid and total crime number.



To check the distribution of our hypothesis test, choose one of the bottom three bottoms as their tile described. As you press the bottom, a corresponding p-value of this test will show on the right in the P-value box. The null hypothesis of each test is stated on top of the bottom. The below screenshot is an example of the first bottom, which is the mean difference of number of total crime.



- Note that each time you press the three bottom for the hypothesis test, the program will generate a new randomization test.
- Note that the p-value for our test might always be 0 or close to 0, this is normal as this is the true result we get from our program.
- Note that please close the graph window before opening a new one, as the graph will overlap if you don't close the existing graph window.

Description of changes

- We changed some of our introduction to make it more precise and concise.
- We removed some of the sub-questions that are similar to avoid redundant analysis.
- We added randomization test to our computation to increase the overall credibility and complexity of our coding.
- We add interactive windows to help conclude all our results and provide clearer way for viewers to access them.

Discussion

In our following discussion, we consider the start of the covid pandemic as January 2020, which is the 11th month after the start of our data. Therefore we mainly discuss the difference between the period before and after 2020.1.

How has pandemic impact unemployment rate?

Based on the linear regression model visualization, there is a gentle increase of the unemployment rate. There was a spike in unemployment rate around 14 months after the start of the covid. Before covid, unemployment rate

floated around 6 percent, and unemployment rate surged to 14 percent after a few months from covid, which fits the reality as covid has caused tons of job losses due to undesirable economic condition and government regulations. Unemployment started to decrease a year after the covid outbreak, as measures are implemented to restore the economy and increase job opportunities. Although decreasing, the post-covid unemployment rate stabilizes around 8 percent, higher than pre-covid condition. This gentle increase is supported by the hypothesis test as the p-value is 0, means that we have very strong evidence against our null hypothesis that there is no difference in the mean unemployment rate with and without covid.

How has pandemic impact the number of crimes committed?

Based on the linear regression model visualization, there is an overall decrease of the number of crimes. We can separate the number of crimes into two sections: before and after the covid outbreak. Around a few months after the start of the covid(2020.1), the number of crimes committed drops and floats around a lower level of 45000 comparing with the number of pre-covid section of around 55000. This suggests that the pandemic substantially lower the number of crimes committed. This decrease is supported by the hypothesis test as the p-value is 0, which means that we have very strong evidence against our null hypothesis that there is no difference in mean total number crimes with and without covid.

How has pandemic impact the number of crimes related to money committed?

Based on the visualization of linear regression model, there is overall decrease of the number of money related crime. We discovered a sudden decrease in the number of money related crime a few months after the start of the covid. Before covid, money related crime committed floated around 17000, then number suddenly decreased to as low as nearly 0 since covid. Next, the number climbed and stabilized around 12000, which is lower than the numbers before covid. This is a relatively unexpected finding as we originally considered that a decreasing economy and increase un-employment would cause more money related crime for the reason that people's need for money intensify. However, the sudden decrease in these crime states the other way, which might be explain by the sudden decrease in people's outing during the pandemic. This decrease is also supported by the hypothesis test as the p-value is 0, which means that we have very strong evidence against our null hypothesis that there is no difference in mean total number of money related crimes with and without covid.

Is there a relationship between unemployment rate and total crime committed?

Based on the visualization of linear regression model, the overall relationship between these two factors is that crime number decrease as unemployment rate increase. During the pre-covid period, the crime number is highest with unemployment rate around 6 percent . During the post-covid period, the crime number is at a much lower level with an unemployment rate round 8 percent.

Is there a relationship between unemployment rate and money-related crime number?

Based on the linear regression visualization, money-related crime number decrease as unemployment rate increase.

Limitation

The crime data we obtain only contains data from March 2019 to August 2021, which suggests that we have only 9 months of data representing the pre-covid situation. This might not be enough to create a generalized relationship between unemployment and crime. The drop in crime might be due to the specialty of a pandemic, as human interaction is substantial reduced due health regulations. Therefore, during the covid period, we might in fact be experiencing a high risk of crime outburst, but these potential crimes are prevented by regulations like quarantine, and if the unemployment rate remains high after the covid era, we might face a real burst of crime commitment.

In conclusion, the covid outbreak have impacted both unemployment rate and the crime committed, and the impacts are especially profound at the start of the pandemic. Unemployment rate increase after the pandemic, and the number of crimes, both in total and specify to money related, decrease after the outbreak. We can conclude a negative relationship between crime and unemployment based on our existing data, but as mentioned in the limitation, future studies are suggested with more pre-covid data to generalize this relation.

References

Canada Unemployment Rate | 2021 Data | 2022 Forecast | 1966-2020 Historical | Calendar. (n.d.). Retrieved November 4, 2021, from <https://tradingeconomics.com/canada/unemployment-rate>

Statistical relationships between crime trends and major socio-economic trends: Findings. (n.d.). Retrieved November 4, 2021, from <https://www150.statcan.gc.ca/n1/pub/85-561-m/2005005/findings-resultats/4144302-eng.htm>

Government of Canada, S. C. (2021, October 8). *Employment and unemployment rate, monthly, unadjusted for seasonality.* <https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1410037401>

Government of Canada, S. C. (2021b, October 22). *Selected police-reported crime and calls for service during the COVID-19 pandemic.* <https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=3510016901>