**CIND820 – Capstone Project**

**Abstract**

**Katie Schilling**

**501130072**

**Supervisor: TBD**

**January 18, 2022**

**Vital Events Predicting, Forecasting and Introduction of Possible Causality**

Every year the registrar general collects and releases information on vital events in Canada. Based on research, the vital events data is collected to monitor, and predict future population growth, marriages, deaths and stillborn births. *(https://www.statcan.gc.ca/en/about/relevant/vscc/faq)*

Much of this information can be used to determine what resources will be needed in the future, whether the population is increasing, decreasing, or holding steady, and foresee any public health trends emerging, as we have seen over the past 2 years with the introduction of COVID-19. All of this information is important for understanding and maintain balance of our system as a whole.

In addition to focusing on the births, deaths, marriages and stillborn births, I also want to introduce the number of positive COVID-19 cases, month over month. The reason for this is to see if there is a direct correlation between the increase or decrease of these numbers and COVID-19, and what long term impact it may have on our growth or decline.

**Goals of the analysis:**

Predict the future population, and trends. Will we see an overall increase in population, or will there be a decrease?

Does the introduction of COVID-19 change the population forecast? Does COVID-19 create an increase of deaths and stillbirths? What impact on marriages and new births does COVID-19 potentially have?

Are we seeing a trend over time of a decrease in stillborn births and deaths as our healthcare system advances, when COVID-19 is not a factor?

Based on history, what numbers did we expect to see in 2020 and 2021? How did those numbers differ with the introduction of COVID-19?

**Tools and Techniques**

R-Studio

Scatter, bar, and box plots, histograms, to visualize the time-series trends

Correlation matrix

Causality test in R

Normalization of the data

Training and test data sets for model validation

Naïve Forecasting Method

Forecasting and prediction

**Data Information**

The dataset being used for the vitals information can be found at:

<https://open.canada.ca/data/en/dataset/e100b1d5-b7fb-40fd-b7a0-50289d13c574>

In order to introduce COVID-19 positive cases, to infer causality from this data, I need to pull data from a secondary source, and match the data to the existing dataset. Initially this proved to be a difficult task, as much of the information on COVID-19 is not publicly available or the raw data is coded, and unclear. After research I found the following data:

https://health-infobase.canada.ca/covid-19/epidemiological-summary-covid-19-cases.html

While the timeline on the 2 data sets does not correlate as is, through pre-processing of the data, I will be able to use the totals, per month, to match the data to the vitals event data. The only data required from this dataset will be prname (province name), to which I will only be using the ones listed under Canada as it encompasses all provinces, and the column numtoday (number today) which is the number of new covid cases since previous update. The month of January, 2020 will be excluded as it is an incomplete month and will skew the results.